Fall 2014

New experimental and theoretical tools for studying protein systems with elements of structural disorder

Tairan Yuwen
Purdue University

Graduate School Form 30
(Updated 11/20/2014)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Tairan Yuwen

Entitled
New Experimental and Theoretical Tools for Studying Protein Systems with Elements of Structural Disorder

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Nikolai R. Skrynnikov

John S. Harwood

Carol B. Post

Alan M. Friedman

Timothy S. Zwier

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Nikolai R. Skrynnikov

Approved by Major Professor(s): _____

_____

Approved by: Christine A. Hrycyna 12/01/2014

Head of the Department Graduate Program Date

NEW EXPERIMENTAL AND THEORETICAL TOOLS FOR STUDYING PROTEIN SYSTEMS WITH ELEMENTS OF STRUCTURAL DISORDER

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Tairan Yuwen

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2014

Purdue University

West Lafayette, Indiana

This work is dedicated to my parents and my grandparents.

ACKNOWLEDGMENTS

First of all I want to express the deepest thank to my major advisor Nikolai Skrynnikov, who has guided me into the fascinating research field of protein NMR. Nikolai is one of the few professors who would like to discuss with his students and provide any kind of help as much as he can. It is impossible for me to make significant achievements during my PhD study without his support or encouragement. He is not only knowledgeable in NMR related research fields but also good at teaching other people, until now I still feel those classes he offered are among the best ones I have ever taken, from which I have learned quite a lot of useful things related to my current research work. I also want to thank all my committee members, Carol Post, Alan Friedman, Tim Zwier and John Harwood, for keeping me on track of my study and research work by asking me insightful questions or giving me helpful advices in my annual advisory committee meetings.

I also want to thank all my group members who have worked together with me in the past or at present. I want to give especially thank to the previous PhD student and postdoc Yi Xue in my lab, from whom I have learned many different things such as protein expression, NMR experimental design, computer setup and programming. I also want to thank two other postdocs Jose Enrique Herbert-Pucheta and Aleksandr Yakimov. Although both of them did not stay in the lab for a long time, I still learned some useful skills from them such as protein crystallization and lab management. Also I want to thank current graduate students Adam Groves and Kyle Chamberlain in my group, who help to maintain the lab in a good atmosphere, besides, it also helps me to form better understanding about many things I learned before by teaching them those things again. I also would like to thank several people in some other research

groups: Haiyan Chu in Prof. Low's lab has helped me a lot during my PhD study, I have borrowed reagents or instruments from her lab many times and she is always glad to offer anything as I need. Nina Gorenstein in Prof. Post's lab taught me many experimental skills about protein expression, without her help it would be difficult for me to start working on my first research project.

I would like to thank Jerry Hirschinger and Huaping Mo in Purdue NMR Facility, who have helped me many times since I started my NMR Facility teaching assistant work several years ago. I am also grateful for the assistance from Donna Bertram, who has provided much convenience for my teaching assistant work as well as many other stuffs related to my lab.

Finally, I would like to express deepest appreciation to my parents, who have supported my PhD study at Purdue during the past several years. Although they both live in China and are thousands of miles away from me, I still feel just staying together with them when I make phone call to them every week. Without their firm support it is impossible for me to finish writing this thesis or complete my PhD study.

TABLE OF CONTENTS

Page

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

SYMBOLS

R_1	spin–lattice (longitudinal) relaxation rate
R_2	spin–spin (transverse) relaxation rate
$R_{1\rho}$	spin–lattice relaxation rate in the rotating frame
R_2^{app}	apparent R_2 rate measured from PD-CPMG
^{15}N–$\{^1H\}$	saturation transfer NOE on ^{15}N
η_{xy}	transverse DD/CSA cross-correlated relaxation rate
k_{ex}	chemical exchange rate constant
R_{ex}	relaxation rate caused by chemical exchange
$\hat{H}_{DD}$	dipole–dipole (DD) interaction Hamiltonian
$\hat{H}_{DD}^{sec}$	secular part of dipole–dipole (DD) interaction Hamiltonian
$\mathbf{A}$ ($\mathbf{S}$)	alignment tensor (Saupe matrix)
$\hat{\mathbf{I}}$ ($\hat{\mathbf{S}}$)	spin angular momentum operator
$\mathbf{B}_0$	external magnetic field vector
$\hat{\rho}$	spin density operator
$\hat{\hat{L}}$	Liouville superoperator (Liouvillian)
$\hat{\hat{R}}$	relaxation superoperator
S^2	order parameter from model-free analysis
$\Delta(^2H)$	quadrupolar splitting in ^2H NMR spectra
$\Delta\delta$	binding shift caused by chemical shift perturbation
$g(\tau)$	correlation function
$J(\omega)$	spectral density function
R_g	radius of gyration
K_d	dissociation equilibrium constant

k_{on}	association rate constant
k_{off}	dissociation rate constant
τ_c	correlation time of stochastic interactions
τ_R	rotational tumbling motion correlation time
τ_{rel}	duration of the relaxation delay in PD-CPMG
τ_{cp}	half of the interval between ^{15}N CPMG pulses in PD-CPMG
τ_{mix}	duration of the CP mixing time in CP-HISQC
ν_1	*rf* pulse field strength (unit in Hz)

ABBREVIATIONS

α-spc SH3	α-spectrin SH3 domain
C12E5	pentaethylene glycol monododecyl ether
CaM	calmodulin
CD	circular dichroism
CP	cross-polarization
CP-HISQC	cross-polarization assisted HISQC
CPMG	Carr–Purcell–Meiboom–Gill
cryo-EM	cryo-electron microscopy
CSA	chemical shift anisotropy
DD	dipole–dipole
DHPC	dihexanoyl phosphatidylcholine
DLS	dynamic light scattering
DMPC	dimyristoyl phosphatidylcholine
Drk	downstream of receptor tyrosine kinase
drkN SH3	N-terminal SH3 domain of *Drosophila* Drk
DSC	differential scanning microcalorimetry
FRET	fluorescence resonance energy transfer
GB1	protein G B1 domain
GDO	generalized degree of order
GPU	graphics processing unit
HISQC	heteronuclear in-phase single quantum coherence spectroscopy
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein

IDR	intrinsically disordered region
INEPT	insensitive nuclei enhanced by polarization transfer
ITC	isothermal titration calorimetry
LJ	Lennard–Jones
MAS	magic angle spinning
MD	molecular dynamics
MDEC	molecular dynamics in electronic continuum
NBD	nucleotide-binding domain
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NSP	normalized scalar product
NSR	nuclear spin relaxation
PCS	pseudocontact shift
PD-CPMG	proton-decoupled CPMG
PDB	Protein Data Bank
PEG	polyethylene glycol
PRE	paramagnetic relaxation enhancement
RDC	residual dipolar coupling
rf	radio-frequency
SAXS	small-angle X-ray scattering
SH2	Src homology 2
SH3	Src homology 3
SPR	surface plasmon resonance
SSNMR	solid-state NMR
Syk	spleen tyrosine kinase
UV	ultraviolet

ABSTRACT

Yuwen, Tairan. PhD, Purdue University, December 2014. New Experimental and Theoretical Tools for Studying Protein Systems with Elements of Structural Disorder. Major Professor: Nikolai R. Skrynnikov.

Disordered proteins are one class of proteins which do not possess well-folded three-dimensional structures as their native conformations. Many eukaryotic proteins have been found to be fully disordered or contain certain disordered regions. Disordered proteins usually display several characteristic properties, such as increased motional freedom and the conformational heterogeneity caused by that. The elements of structural disorder are commonly involved in many important biological functions and are implicated in many diseases. Therefore, the study of disordered proteins has become one of the most important research topics in recent years. This thesis presents results from three different research projects; the common feature is that all systems being studied contain varying amount of structural disorder. Most results have been obtained based on experimental nuclear magnetic resonance (NMR) studies and molecular dynamics (MD) simulations. Both are among the most popular biophysical techniques for studying molecular dynamics. The first project investigates the relationship between domain cooperativity and residual dipolar coupling (RDC) parameters based on a series of two-domain chimera proteins with disordered linkers. Many eukaryotic proteins contain multiple domains and their biological functions are closely related to the property of domain cooperativity, which is often regulated by the linker region. Therefore it is necessary to develop suitable tools to characterize linker region properties in order to better understand biological functions of multidomain proteins. The second project is about the development of NMR pulse sequences for studying

disordered proteins. Two new NMR pulse sequences, PD-CPMG and CP-HISQC, have been developed. Both experiments are well suited for studying intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) under physiological conditions. These two experiments produce higher precision for ^{15}N R_2 rates measurement or higher sensitivity in ^{1}H–^{15}N HSQC spectra respectively. Besides, they also show many advantages over most other existing experiments for studying IDPs. The last project is about protein–peptide encounter complex study based on Crk–Sos model system. The ten-residue Sos peptide serves as a minimal model for disordered proteins. Encounter complex is an important type of intermediate state formed during many protein interactions. Such complexes are usually characterized by a large amount of motional freedom and conformational heterogeneity. Therefore their properties are considerably different from tight-binding complexes which are more commonly studied. Although it is usually quite difficult to study encounter complexes using standard biophysical techniques, in this project we have successfully characterized structural and dynamic properties of Crk–Sos electrostatic encounter complex with a combination of MD simulations and experimental NMR approaches. It can be directly seen from the structural model based on MD trajectories that Sos peptide in the encounter complex remains highly dynamic, sampling large area on the surface of Crk N-SH3 domain. Such strategy can also be utilized for studying many other encounter complexes involving disordered proteins or peptides.

CHAPTER 1. INTRODUCTION

1.1 Domain Cooperativity of Multidomain Proteins Investigated through Residual Dipolar Couplings (RDCs)

1.1.1 The Role of Linker Regions in Multidomain Proteins

Proteins are the basic building blocks of life. The diversified biological functions of different proteins are associated with different structural or conformational properties, which in turn are determined by the unique amino acid sequence of each protein according to Anfinsen's dogma[1]. Most proteins (except for those that are fully disordered) contain different numbers of structural units called domains. According to the genomics analyses, more than 70% of eukaryotic proteins contain multiple domains[2]; this proportion is much higher than the corresponding ratio for prokaryotic proteins. Multidomain proteins are commonly found to perform more complicated biological functions than single-domain proteins. This is a good example that illustrates the basic principles of biological evolution. For multidomain proteins, not only the structure of each individual domain but also the relationship between different domains are important for carrying out their proper biological functions.

The modular architecture of multidomain proteins is commonly observed in those proteins that are involved in signaling transduction pathways. Such proteins typically contain several signaling domains (e.g., SH2 and SH3 domains) connected via interdomain linker regions[3]. Since each signaling domain usually behaves as an adaptor and can interact with multiple binding partners, such modular architecture provides an efficient strategy to bring together multiple components in signaling transduction pathways, further leading to meaningful downstream events.

The linker regions within multidomain proteins not only play a role as domain connectors, but also regulate biological functions of the entire protein[4,5] or modulate multiple protein–protein interactions[6]. Many instructive examples can be found that illustrate the importance of linker regions within multidomain proteins[7]. Calmodulin (CaM) contains two domains connected by the linker region. The linker region is relatively rigid in the absence of Ca^{2+}, but becomes much flexible when Ca^{2+} binds to this protein. Therefore calmodulin exists in two different conformational states characterized by a different degree of domain–domain coupling as determined by the linker region. These two states are associated with different biological functions[8]. Protein tyrosine kinase Syk has tandem SH2 (tSH2) domain architecture. The two SH2 domains are connected by the 45-residue linker region which is relatively rigid in the unphosphorylated form; however, it becomes much more flexible once the linker region residue Y130 is phosphorylated. This event directly causes the activation of Syk kinase, leading to many further downstream effects[9].

One of the common features that can be surmised from these examples is that multidomain proteins can exist in multiple states characterized by a different amount of domain–domain coupling. The latter is often determined by the property of linker regions—rigid or very short linkers correspond to fully structurally coupled domains, whereas long and flexible linkers correspond to uncoupled domains[1]. Since structural and dynamic properties of linker regions can be modulated by many different factors (such as Ca^{2+} binding or phosphorylation), the degree of domain–domain coupling can also be modulated in many different ways, which forms the basis of "conformational switch" mechanism in many multidomain signaling proteins[10]. Due to the important biological role of linker regions within multidomain proteins, it is essential to develop suitable tools to characterize linker region properties in order to better understand the relationship between domain cooperativity and biological functions of multidomain proteins.

[1]Assume that there are no direct interactions between different domains.

1.1.2 Biophysical Study of Multidomain Proteins

The structure/dynamics characterization of multidomain proteins is much more challenging than the study of single-domain proteins, mainly due to the conformational heterogeneity caused by flexible linkers[11,12]. If the linker region within multidomain protein is totally rigid, then the whole protein should behave as a rigid body and can be treated similar to a single-domain protein. However, in many cases linker regions display a significant amount of motional freedom (structural disorder), such that multidomain proteins cannot be adequately represented by one single conformation. Instead, it is necessary to resort to ensemble-based approaches to fully characterize the properties of such proteins[13–15].

Many different experimental techniques have been developed for studying protein structure and dynamics. X-ray crystallography is one of the most powerful tools for obtaining structural information. However, it encounters significant difficulties when studying multidomain proteins: the structural information obtained from X-ray crystallography is usually limited to one single conformation[16], such that it is not ideally suited for studying conformationally heterogeneous systems. In spite of that, certain useful information (such as the structure of each individual domain) can still be obtained from X-ray crystallography, which can be further used as an input for other biophysical methods such as SAXS and cryo-EM[17]. Among various experimental techniques, NMR and SAXS are two most popular methods for studying multidomain proteins.

Nuclear magnetic resonance (NMR): The most commonly used NMR tools for studying multidomain proteins include nuclear spin relaxation (NSR) and RDCs[11,12]. Protein backbone ^{15}N relaxation parameters are commonly used to probe molecular picosecond–nanosecond timescale dynamics. One of the popular strategies for ^{15}N relaxation data analysis is the Lipari–Szabo model-free method[18], which is based on the assumption that fast internal motion is independent from the overall molecular tumbling. The rotational diffusion tensor of each single domain within a multidomain protein can be obtained from such analysis; then the property of linker regions can

be inferred based on the relationship between the diffusion tensors from different domains. RDC-based studies follow a similar path, except that the role of rotational diffusion tensor is now played by the alignment tensor. There have been a significant number of studies about multidomain proteins based on the joint analysis of nuclear spin relaxation data and RDCs[19–22]. Besides, several other types of NMR parameters such as paramagnetic relaxation enhancement (PRE) and pseudocontact shift (PCS) have also been utilized for studying multidomain proteins[23,24].

Small-angle X-ray scattering (SAXS): SAXS studies are mainly based on the analysis of scattering intensity profile $I(s)$ measured in solution-state samples. Different aspects of information about molecular structure such as overall shape and radius of gyration R_g can be obtained from such analysis[25]. SAXS studies are commonly used in combination with structural modeling approaches, generating structural ensembles to reflect the conformational heterogeneity of biomolecules[26]. There have also been a number of studies based on the joint analysis of SAXS and NMR data[27–29].

Aside from SAXS and NMR, several other biophysical techniques have also been utilized for studying multidomain proteins. Fluorescence resonance energy transfer (FRET) can report on domain–domain distance based on the relationship of transfer efficiency and the separation between FRET donor–acceptor pairs. Therefore, the spatial relationship between domains can be derived based on such analyses. Most traditional FRET studies report on ensemble-averaged parameters. In contrast, single-molecule FRET makes it possible to obtain certain information pertaining to the individual molecules[30]. Similar to SAXS, cryo-EM can provide low-resolution structural information (such as the overall shape), which is also commonly used in combination with structural modeling methods to characterize multidomain proteins[31].

1.1.3 Theoretical Backgrounds of RDCs

Residual dipolar coupling (RDC) effect has a long history in the field of NMR spectroscopy. The first study of RDCs was carried out by Saupe and Englert on small organic molecules in a liquid crystal[32]. In their pioneering work they proposed

some key concepts pertaining to interpretation of RDCs, such as alignment tensor $\mathbf{A}$ (or Saupe matrix $\mathbf{S}$) which characterizes molecular orientation anisotropy in liquid-crystalline solvent. RDCs arise from the incomplete averaging of dipole–dipole (DD) interactions between a pair of nuclei (considering spins I and S). The effects of DD interaction can be described by dipolar Hamiltonian ($\hat{H}_{DD}$), which can be expressed as:

$$\hat{H}_{DD} = -\frac{\gamma_I \gamma_S \mu_0 \hbar}{4\pi^2 r_{IS}^3} \left(3\hat{\mathbf{I}} \cdot \frac{\mathbf{r}_{IS}\mathbf{r}_{IS}}{r_{IS}^2} \cdot \hat{\mathbf{S}} - \hat{\mathbf{I}} \cdot \hat{\mathbf{S}} \right) = -\frac{\gamma_I \gamma_S \mu_0 \hbar}{4\pi^2 r_{IS}^3} \hat{\mathbf{I}} \cdot \mathbf{D} \cdot \hat{\mathbf{S}} \tag{1.1}$$

where γ_I and γ_S are gyromagnetic ratios of spins I and S, μ_0 is the magnetic permittivity of free space, $\mathbf{r}_{IS}$ is the vector connecting I and S and r_{IS} is the respective distance, $\hat{\mathbf{I}}$ and $\hat{\mathbf{S}}$ are spin angular momentum operators for spins I and S, $\mathbf{D}$ is the dipolar tensor. Under high-field approximation, only the secular part of $\hat{H}_{DD}$ remains important, such that it can be expressed in a truncated form (θ_{IS} is the angle formed by $\mathbf{r}_{IS}$ and external magnetic field $\mathbf{B}_0$):

$$\hat{H}_{DD}^{sec} = -\frac{\gamma_I \gamma_S \mu_0 \hbar}{4\pi^2 r_{IS}^3} (3\cos^2\theta_{IS} - 1) \hat{I}_z \hat{S}_z \tag{1.2}$$

The dipolar interactions shift the energy levels of nuclear spin eigenstates (i.e., Zeeman states), which in turn causes shifts in the position of spectral peaks in NMR spectra. The effects of DD interactions are most commonly observed in solid-state NMR (SSNMR). Each molecule within solid powder samples has a fixed orientation and thus $\mathbf{r}_{IS}$ is fixed[2], when considering the ensemble of molecules within each sample, the distribution of θ_{IS} ($\in [-\pi, \pi]$) leads to familiar powder pattern spectra[33].

The effective magnitude of dipolar interaction is also affected by many other factors such as motion of the sample. The most important examples are magic angle spinning (MAS) in solid-state NMR and overall molecular tumbling in solution-state NMR. In those two cases, the time-averaged value of $\hat{H}_{DD}^{sec}$ is evaluated as:

$$\langle \hat{H}_{DD}^{sec} \rangle = -\frac{\gamma_I \gamma_S \mu_0 \hbar}{4\pi^2 r_{IS}^3} \hat{I}_z \hat{S}_z \langle 3\cos^2\theta_{IS} - 1 \rangle \tag{1.3}$$

[2]Assume for a moment that the distance r_{IS} is fixed.

The MAS works under the "magic angle" condition such that $\langle 3\cos^2\theta_{IS} - 1\rangle = 0$, which yields $\langle \hat{H}_{DD}^{sec}\rangle = 0$; thus severe line-broadening effects in SSNMR spectra can be largely suppressed. In solution, isotropic molecular tumbling provides a natural mechanism to average out $\hat{H}_{DD}$; therefore DD interactions only manifest themselves through second-order effects such as nuclear spin relaxation[34]. However, in certain cases molecular tumbling motion is not completely isotropic, such that $\langle \hat{H}_{DD}^{sec}\rangle \neq 0$ and there still exists a residual amount of dipolar coupling termed RDC. In protein NMR spectroscopy, such situation is usually created by adding an alignment media to the sample or by introducing a paramagnetic tag, which induces molecular orientation anisotropy.

1.1.4 Applications of RDCs for Biomolecular Studies

Although RDCs were discovered more than half of a century ago, their applications to biomolecular studies have a relatively short history. RDCs started to become popular only after several types of alignment media compatible with biomolecules have been introduced[35]. The alignment media commonly used for studying biomolecules include DMPC/DHPC[36], phage (Pf1)[37] and PEG/hexanol[38]; besides, several other types of alignment media such as cellulose nanocrystals[39] and DNA nanotubes[40] have also been introduced recently. Currently RDCs have become one of the most important NMR tools in the area of biomolecular structure and dynamics studies[41–45].

Prior to the advent of RDCs, the information about protein structure was mainly obtained from ^{1}H–^{1}H NOEs[46]. However, the fact that NOE intensities are proportional to r^{-6} implies that NOEs can only provide short-range ^{1}H–^{1}H distance information (< 5 Å). This constitutes a significant limitation for protein structure determination based solely on NOEs. Since the magnitude of RDCs are mainly determined by the term $\langle 3\cos^2\theta_{IS} - 1\rangle$ according to Equation (1.3), RDCs are mainly used as angular restraints for protein structure determination, complementary to distance restraints provided by NOEs. In recent years, many types of paramagnetic ions (such as Ln^{3+}) have been used to achieve protein alignment in external magnetic field,

which makes it possible to measure RDCs as well as several other parameters such as PREs and PCSs on the same protein sample[47]. These data can be combined and used together as an input for protein structure determination, for which the support is offered in many programs such as Xplor-NIH[48] and CS-Rosetta[49]. More accurate protein structure information can be obtained from such joint data analyses than based on single type of NMR data only.

RDCs also report on molecular dynamics. Compared with nuclear spin relaxation, the dynamics information obtained from RDCs does not suffer from the limitation imposed by overall molecular tumbling. Therefore, in principle it can report on molecular dynamics in a much larger time window (up to ~ms timescale)[50]. Protein motions happening on relatively slow timescales can be identified by comparing RDC-based order parameters S^2_{RDC} with Lipari–Szabo order parameters S^2_{LS}. The microsecond–millisecond time scale motions probed in this manner (termed supra-τ motions) are relevant to many important biological functions[51,52]. The main setback of RDC-based dynamics analysis is that it requires an extensive set of RDC data measured in multiple alignment media "orthogonal" to each other (i.e., sufficiently different in terms of alignment parameters), therefore it is more time-consuming than many other types of studies in the area of protein dynamics[53].

1.1.5 Summary of Original Contributions

In this study, we have constructed a series of two-domain proteins with different lengths of disordered linker regions to investigate the relationship between domain cooperativity and RDC alignment parameters[54]. The main idea is that two-domain proteins with very short linkers should behave similar to a rigid body due to highly restricted domain–domain dynamics, whereas two domains connected by a long and flexible linker should behave as two independent entities. It is expected that a gradual transition should exist between these two limiting cases such that degree of domain cooperativity can be fine-tuned by adjusting the length of disordered linker. Previous studies with similar ideas have been carried out on certain model systems (such

as GB1[22]); however, these studies focused on nuclear spin relaxation parameters instead of RDCs. In this study we focus on using RDC parameters to characterize domain cooperativity of multidomain proteins. Towards this goal we have carried out both *in silico* structural modeling and experimental studies seeking to achieve better understanding of multidomain proteins.

The *in silico* structural modeling studies have been carried out for a series of two-domain proteins with different domain sizes. Each two-domain protein construct is represented by an ensemble of conformations, with each conformer containing two identical domains connected by a flexible linker region. The linkers have $(GGSGG)_n$-type sequence, motivated by the fact that Gly and Ser possess the largest amount of conformational flexibility among all amino acids (Ser is added to have a polar side chain, mainly for the sake of solubility)[55,56]. The program TraDES[57] is used to generate flexible linkers, which are in turn used to "glue together" the two domains. Each ensemble originally contains ~ 10000 structures, then each conformer within the ensemble is subjected to 1 ps torsional angle dynamics refinement using the program Xplor[48]. After this step we only retain those structures that have relatively low free energies to construct the final ensemble, which usually consists of 2000–3000 conformers. Such strategy is supported by the results from random energy model (REM), which describes the free energy distribution of random polymers[58].

For each two-domain construct, RDC parameters are predicted by the program PALES[59] using simulation parameters mimicking the experimental conditions. The PALES calculations have been carried out for each individual conformer within the ensemble, then RDCs were averaged over all conformers and, finally, also averaged over N- and C-terminal domains (consistent with the experimental results from tandem α-spc SH3 domains). Several RDC-based parameters (such as GDO and NSP) show clear dependence on the length of linker regions. Furthermore, such dependence has been observed for several two-domain proteins with different domain sizes. These results can be explained in such a way: the two-domain construct with a short linker is highly elongated such that it tends to produce large steric alignment, whereas in the

case of a long and flexible linker it should produce minimum steric alignment because each individual domain is near spherical. Therefore, the results from *in silico* structural modeling strongly support the idea that domain cooperativity in two-domain proteins can be characterized with RDC alignment parameters.

The experimental study focuses on three two-domain constructs based on α-spc SH3 domain. Specifically, those three chimera proteins have identical N- and C-terminals. These domains are connected by linker regions with different lengths, ranging from 4 to 30 residues. For each two-domain protein, we have prepared ^{15}N-labeled samples to measure backbone amide $^{1}H^{N}$–^{15}N RDCs. PEG/hexanol alignment media has been chosen, since this alignment media is stable over a wide range of sample conditions and it offers the alignment mechanism that is mainly steric[38]. RDCs are measured using IPAP-HSQC scheme[60], alignment tensor **A** is extracted from the fitting of RDCs to high-resolution X-ray structure of α-spc SH3 domain (PDB code 1U06[61]). We have analyzed RDC parameters as obtained for those three tandem proteins and found that our experimental results show relatively good agreement with the trend predicted by *in silico* structural modeling. However, we have also observed certain discrepancies between experimental data and *in silico* structural modeling, especially for the single-domain α-spc SH3. We have investigated potential reasons for such discrepancy and proposed possible explanations. This study provides general guidelines for studying multidomain proteins via RDC alignment parameters. Our results should be helpful for similar future studies of biologically relevant multidomain proteins.

1.2 NMR Pulse Sequence Development for Disordered Protein Studies

1.2.1 Overview of Basic Properties and Biophysical Studies about Disordered Proteins

According to traditional structure–function paradigm, each protein must possess specific three-dimensional structure in order to carry out its unique biological functions, however, in recent years this paradigm has received much challenge due the

discovery of disordered proteins. Disordered proteins also carry out many important biological functions but they do not possess well-folded structures as their native conformations[62]. Such proteins are most prevalent in higher level organisms: more than 50% of proteins in eukaryotic cells are fully disordered (intrinsically disordered proteins, IDPs) or contain certain disordered regions (intrinsically disordered regions, IDRs)[63]. The study about disordered proteins has become one of the most important research topics due to their important biological functions and the relationship with different types of pathologies[64,65]. The most well-known examples include α-synuclein and Aβ, both are associated with neurodegenerative diseases and have been extensively studied via simulation and experimental based approaches[66]. IDPs are fully disordered within the whole amino acid sequence, whereas IDRs are usually located at protein terminal regions to carry out important biological functions, for example, the N-terminal disordered tail of human p53 is crucial for its DNA binding property[67].

One significant feature of disordered proteins is the promiscuous binding property caused by their structural plasticity[68]. IDPs usually play important roles as the hubs of protein interaction networks, especially in cell signaling or cell regulation pathways[69,70]. Since IDPs do not require folded part as scaffold to support their biological functions, the disordered structural nature is usually considered to be more economical and functional advantageous than their folded protein counterparts[71–73].

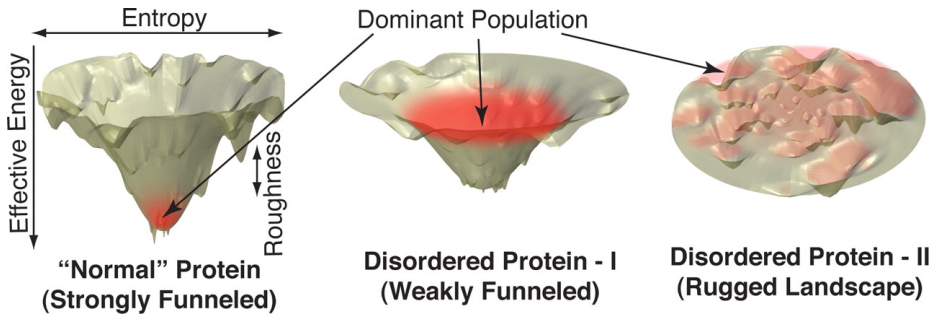


Figure 1.1. Typical forms of potential energy landscapes for folded and disordered proteins[74].

The structural feature of IDPs can be well understood by the aid of potential energy landscape theory[58]. The free energy landscape of folded proteins usually contains one single energy minimum corresponding to the natively folded state (Figure 1.1A), besides, the native conformation usually occupies $\sim$ 100% population since the potential energy within this basin is much lower than other regions. Disordered proteins usually have much different potential energy landscapes, which are either weakly funneled or highly rugged (Figure 1.1B,C)[74]. Such free energy landscape allows multiple conformations interconverting with each other on many different timescales[75], therefore structural features of IDPs can only be completely characterized with ensemble-based approaches[76,77].

The distinct structural or conformational properties of IDPs can be traced down to their characteristic amino acid composition. Most disordered proteins are enriched in polar (such as Gln or Ser) or charged residues but lack hydrophobic residues, therefore it cannot form the hydrophobic core as required by most folded proteins[78]. The relationship between amino acid composition and structural disorder forms the basis of many bioinformatics algorithms to predict IDPs (or IDRs) within genomes of different organisms[79–81]. Besides, many databases about disordered proteins have been constructed (such as DisProt[82]), which allows the identification of IDPs (or IDRs) based on the amino acid sequence more easily.

The structural and dynamic properties of disordered proteins can be studied with different types of biophysical techniques. The disordered structural nature indicates that IDPs are difficult to crystallize such that it is not suitable for X-ray crystallography study, instead some other biophysical methods such as NMR and SAXS are better suited for studying IDPs. NMR is usually considered as the most powerful tool in IDP related studies, more details will be presented in later sections. SAXS is commonly used in combination with different structural modeling approaches. Ensemble Optimization Method (EOM)[83] is one of the most popular algorithms, which selects a small number of structures from a large conformer pool that produce best fit to experimental data (such as scattering intensity profile $I(s)$)[84], hence the con-

formational heterogeneity of IDPs is represented by the variation of structures within the ensemble. There are quite a few studies based on the joint analysis of SAXS and NMR data[85], besides, single-molecule FRET[86], differential scanning microcalorimetry (DSC)[87], dynamic light scattering (DLS)[88], size-exclusion chromatography[89] and circular dichroism (CD)[90] have also been utilized for studying IDPs.

Although disordered proteins possess large amount of structural disorder, the structural features of IDPs should still be distinguished from random coils—disordered proteins usually contain certain amount of residual secondary or tertiary structures which are crucial for their biological functions[91,92]. Such residual structures and their interactions are usually transient in nature, therefore it is difficult to study them with most standard biophysical techniques, however, such transient structures can still be well captured by several NMR-based approaches, such as PRE and RDC[93,94].

1.2.2 NMR Study of Disordered Proteins

Although NMR is usually considered as the most powerful tool for studying IDPs, there are much more challenges than studying folded proteins due to several specific properties of IDPs[95]: (i) The chemical shifts of most nuclei usually have narrow dispersions due to the structural disorder, which easily cause peak overlap in NMR spectra. (ii) IDPs usually contain multiple amino acids of the same type or prolines, which causes large difficulty to NMR peak assignments. (iii) Most regions of IDPs are well exposed to solvent and thus have low solvent protection factors, therefore studies on labile protons (such as amide group $^{1}H^{N}$) suffer a lot from fast solvent exchange process under physiological conditions, which significantly undermines signal-to-noise of NMR spectra. (iv) The disordered structural nature also indicates that IDPs are susceptible to proteolysis or other types of degradation, therefore NMR samples usually have relatively short lifetime and limited measurement time.

In spite of that, quite a few strategies have been developed to overcome difficulties for NMR studies about IDPs: (i) The peak overlap issue can be largely resolved by recording higher-dimensional NMR spectra, quite a few four- or five-dimensional

NMR experiments have been developed specifically for studying IDPs[96–99]. (ii) One of the major drawbacks for recording higher-dimensional NMR spectra is the ultra-long measurement time cost by such experiments, however, the total measurement time can be much reduced by implementing specially designed sampling schemes such as non-uniform sampling (NUS)[100–102]. (iii) Many NMR peak assignment strategies have been proposed for studying IDPs, which can overcome certain issues such as the presence of prolines[103–105], besides, several automatic peak assignment strategies for IDPs have been developed[106,107]. (iv) The low signal-to-noise in NMR spectra caused by fast solvent exchange can be partially overcome by performing frequent pulsing (such as SOFAST-HMQC[108]), which can still maintain good signal-to-noise under relatively fast solvent exchange conditions ($k_{ex} \sim 10\ s^{-1}$). As solvent exchange rates become very large ($k_{ex} > 100\ s^{-1}$), it is common practice to completely avoid studying those labile protons (such as amide group $^{1}H^{N}$), many heteronuclei detection experiments have been designed based on such idea[109–111]. These strategies can also be combined together to create even better experiments for studying IDPs[112,113].

Since IDPs display large amount of motional freedom and they do not possess well-folded three-dimensional structures, most NMR studies about IDPs focus on their dynamic aspects. Protein dynamics at different timescales can be studied with different types of NMR experiments (Figure 1.2)[50,114]. A brief summary about the most commonly use NMR approaches for studying disordered proteins is presented below.

Nuclear spin relaxation (NSR): The data analysis of nuclear spin relaxation parameters for disordered proteins usually follows quite different strategies from folded protein studies. It is usually difficult to separate fast internal motion from overall tumbling motion for disordered proteins, therefore the basic assumption of Lipari–Szabo model-free analysis breaks down. Some alternative data analysis strategies such as spectral density mapping are more suitable for studying IDPs[115,116].

Residual dipolar couplings (RDCs): RDCs report protein dynamics in a broad range of timescales, besides, it can also detect long-range contacts within IDPs to

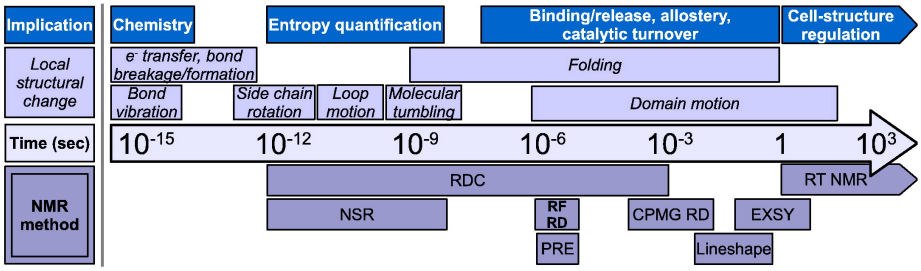


Figure 1.2. NMR methods for studying protein dynamics over a broad range of timescales and the correlation with different biological functions[50]. RT NMR: Real Time NMR; EXSY: EXchange SpectroscopY; CPMG RD: Carr–Purcell–Meiboom–Gill Relaxation Dispersion; RF RD: Rotating Frame Relaxation Dispersion; NSR: Nuclear Spin Relaxation; RDC: Residual Dipolar Coupling; PRE: Paramagnetic Relaxation Enhancement.

provide detailed information about transient structures within IDPs[117]. RDCs are also commonly used as input parameters for structural modeling programs to generate structural ensembles that produce best agreement with experimental data[118,119]. One of the potential issues in RDC-based studies is that alignment media may interact with IDPs and perturb their delicate conformational equilibria, therefore the type of alignment media for RDC measurement should be chosen carefully[95].

Paramagnetic relaxation enhancement (PRE): PREs refer to the enhanced relaxation rates caused by spatially close paramagnetic species. The PRE effect is proportional to r^{-6} and thus behaves similar to NOEs[120], however, the distance information as obtained from PREs covers much larger distance range (up to $\sim$ 40 Å), therefore PREs are suitable for studying biomolecules with extended structures (such as disordered proteins)[121]. PREs directly report long-range order information within IDPs, such as those transient contacts which are closely linked with their biological functions[122]. One of the critical issues for PRE-based studies is the choice of positions for conjugating paramagnetic labels. When disordered protein has relatively long primary sequence, it would be necessary to place paramagnetic labels in several different positions to obtain complete structural information[123]. The introduction

of paramagnetic labels usually requires specific residue mutations, which may also perturb conformational properties of IDPs. PREs are commonly utilized in combination with some other parameters such as RDCs for disordered protein structural modeling[124].

Several other types of NMR parameters have also been utilized for studying IDPs, such as chemical shifts[125] and solvent exchange rates on labile ^{1}H[126]. Besides, some novel techniques have also been developed for studying IDPs, for example, high-pressure NMR provides certain information which cannot be obtained under normal conditions[127]. There have also been quite a few computational tools targeting at disordered protein structure characterization[128], those programs typically generate conformational ensembles based on the input NMR data, such as *flexible-mecanno* and ASTEROIDS[129–131].

1.2.3 Summary of Original Contributions

Although many NMR experiments have been developed specifically for studying IDPs, there are still certain issues remaining unresolved, such as the low signal-to-noise in ^{1}H–^{15}N HSQC spectra caused by fast solvent exchange process. In this work, we have developed two new NMR pulse sequences which mainly target at studying disordered proteins under physiological conditions. These two new pulse sequences display many advantageous features over most other experiments, and they are also suitable for studying folded proteins.

PD-CPMG is developed for ^{15}N R_2 rates measurement for disordered proteins[132], this experiment provides much higher measurement precision than many other commonly used experiments. CP-HISQC records ^{1}H–^{15}N HSQC spectra with much higher spectral resolution than some other popular schemes, besides, it still maintains relatively good signal-to-noise under fast solvent exchange conditions ($k_{ex} \sim 100\ s^{-1}$)[133]. The development of both pulse sequences follows the strategy that combines both numerical simulation and experimental measurement. Numerical simulation provides as general guidelines for NMR experimental design and it also helps to understand

detailed mechanisms of each pulse sequence. On the other hand, it is also essential to obtain experimental data to confirm those results as predicted by numerical simulations. Such strategy has also been utilized for developing many other NMR pulse sequences[134].

For protein ^{15}N R_2 rates measurement, ^{15}N spin-lock[135] and CPMG[136] are two of the most popular approaches, however, both schemes require large amount of *rf* power deposited through ^{15}N channel, therefore there usually exists an upper limit of relaxation delay period τ_{rel} can be used, which is typically 150–200 ms to avoid sample overheating or probe damaging. Since IDPs (or IDRs) display large amount of motional freedom, ^{15}N R_2 rates are usually on the order of 1–2 s^{-1} under physiological temperature (i.e., 37 °C), such that the limit placed on τ_{rel} can easily incur large measurement errors[137]. The main idea of PD-CPMG is to utilize high-quality ^{1}H decoupling (i.e., DIPSI-2[138]) to suppress J-couplings ($^1J_{\mathrm{NH}}$) and DD/CSA cross-correlation. Since ^{1}H has much larger gyromagnetic ratio γ than ^{15}N ($\gamma_{\mathrm{H}} \simeq 10\gamma_{\mathrm{N}}$), the actual *rf* power deposited on the sample should be much lower than spin-lock or CPMG experiments according to the relationship of $\omega_1 = \gamma B_1$, such that much longer τ_{rel} can be sampled with PD-CPMG scheme. The results as obtained from PD-CPMG are fully consistent with some other experiments such as ^{15}N spin-lock, besides, the new scheme provides much higher measurement precision (especially on low concentration protein samples). There are also many other advantages associated with PD-CPMG, for example, ^{15}N offset correction is much more simple than traditional CPMG experiment[139], it is robust to pulse miscalibration and *rf* inhomogeneity, it is also less susceptible to artifacts caused by fast solvent exchange process.

CP-HISQC is mainly developed for recording ^{1}H–^{15}N HSQC spectra for disordered proteins with better signal-to-noise under physiological conditions. ^{1}H–^{15}N HSQC is the most routinely performed experiment for protein NMR study, however, ^{1}H–^{15}N HSQC spectra for IDPs recorded under near physiological conditions suffer a lot from fast solvent exchange process, which significantly undermines signal-to-noise of the spectra. The key component of CP-HISQC is the cross-polarization (CP) element

which conducts the initial step of magnetization transfer (i.e., $^{1}H^{N} \rightarrow {}^{15}N$). In the presence of fast solvent exchange, CP element produces much higher efficiency than many other schemes such as INEPT, which is mainly due to the fact that it effectively makes use of water magnetizations to improve coherence transfer efficiency. The CP-HISQC pulse sequence has also been designed to well preserve water magnetization, which helps to achieve better signal-to-noise in ^{1}H–^{15}N HSQC spectra. CP-HISQC shows much higher sensitivity than the original HISQC experiment[140] under fast solvent exchange conditions ($k_{ex} \sim 100\ s^{-1}$), besides, it achieves much higher spectral resolution in ^{15}N dimension than fast pulsing schemes such as SOFAST-HMQC[108]. We have also demonstrated that CP-based coherence transfer element still remains efficient in the presence of very fast solvent exchange ($k_{ex} \sim 1000\ s^{-1}$), therefore this single element can also be incorporated into many other NMR pulse sequences to help boost signal-to-noise, which include many heteronuclei detection experiments[141]. Besides, since CP-HISQC makes use of ^{15}N in-phase magnetization for ^{15}N frequency labeling, this experiment also serves as a perfect template for designing many other experiments for studying disordered proteins, such as RDCs, PREs or ^{15}N relaxation rates measurement.

1.3 Protein–Peptide Encounter Complex Studied with Crk–Sos Model System

1.3.1 The Role of Electrostatic Encounter Complexes in Protein–Protein or Protein–Ligand Interactions

Protein–protein or protein–ligand interactions play many important biological roles in living organisms, for example, those interactions are commonly found to regulate many other biological events or trigger some further downstream processes[142]. The overall process of protein interactions can be described by different types of binding models[143]. The two-state binding model assumes that there are mainly two states present during the whole process, which correspond to free and bound forms of biomolecules. However, the two-state model is usually oversimplified for describing

most protein interactions due to the large molecular size and complicated free energy landscape of each protein. Aside from free and bound states, many other types of intermediate states could also exist, and encounter complex is one of the most commonly studied intermediate states.

The structural and dynamic properties of encounter complexes have been extensively studied based on several model systems, such as barnase–barstar complex[144] and a series of protein complexes involved in biological electron transfer process[145]. The dynamic properties of encounter complexes are quite different from tight-binding complexes: the ligand is just loosely bound to the protein and still maintains large motional freedom, such that it can perform extensive conformational search until the correct final binding site has been located[146–149]. Encounter complexes are commonly found to exist as ensembles of heterogeneous conformations—if the position of protein within the complex is fixed, the ligand may sample a large amount of surface area rather than at one single position[147]. Most studies about encounter complexes have been carried out based on systems composed of two folded proteins, but much fewer studies have been reported about encounter complexes formed by disordered proteins or peptides[150]. Detailed mechanisms about protein interactions can be learned from encounter complex studies, besides, there are also some other potential applications, for example, many drugs are small peptides and their interactions with protein targets display quite a lot of encounter complex features.

The association kinetics of encounter complex formation have also been well studied based on several model systems, several common features have been observed from those studies[151–153]. The encounter complex formation is usually driven by strong electrostatic interactions between the two binding partners, which can boost association rate constant k_{on} several orders of magnitude over random diffusion process: it is well known that k_{on} is typically $10^5 \sim 10^6$ $M^{-1}s^{-1}$ under the random diffusion limit, whereas it can achieve $k_{on} \sim 10^8$ $M^{-1}s^{-1}$ even up to $\sim 10^9$ $M^{-1}s^{-1}$ due to strong electrostatic attraction forces[154–156]. Such large k_{on} is an essential feature to ensure both fast turnover rates and relatively high binding affinities as required by

certain protein complexes, for example, most photosynthetic redox proteins rely on such properties to carry out their biological functions properly (i.e., electron transfer process)[157]. Since the association step towards forming encounter complexes often relies on strong electrostatic interactions, such intermediate states are often called "electrostatic encounter complexes".

1.3.2 Biophysical Study of Encounter Complexes

Many different biophysical methods have been developed for studying protein–protein or protein–ligand interactions[158,159], besides, several *in silico* approaches such as docking or template-based modeling are also commonly used for studying protein interactions[160,161]. However, there are much fewer studies about encounter complexes than for tight-binding complexes, which is mainly due to several major difficulties for studying such complexes: (i) Encounter complexes usually possess higher free energies than corresponding tight-binding complexes, therefore such intermediate states usually have very low population. Besides, the short lifetime of many encounter complexes also causes large difficulty for most biophysical studies[162,163]. (ii) Encounter complexes with relatively high population have only been found on those protein complexes with very low binding affinities ($K_d \sim 1$ mM), such that it is not suitable to study encounter complexes with certain biophysical methods such as ITC or SPR[164,165]. (iii) Encounter complexes usually display significant conformational heterogeneity caused by the highly dynamic property, therefore such complexes cannot be well represented as one single conformation and thus not suitable for X-ray crystallography study.

NMR is one of the most powerful tools for studying protein–protein or protein–ligand interactions[166–168], it can also be used to study protein complexes with very low binding affinities[169]. Among different types of NMR-based approaches, PRE is the most popular method for studying encounter complexes[120,170], which is mainly due to the fast that PREs are very sensitive to low population states (as low as $\sim 0.5\%$)[171,172]. However, structural models constructed based on PRE data are usually composed of

ensembles of static structures, which cannot fully capture the real dynamic behaviors of encounter complexes.

Aside from PRE, several other NMR-based methods are also commonly used for studying encounter complexes. Relaxation dispersion (RD) is another commonly used approach targeting at the study of low population states (such as those so-called "invisible states")[173], and it has been successfully applied for studying several protein folding intermediates[174]. The successful application of relaxation dispersion usually requires relatively large chemical shift difference ($\Delta\omega$) between interconverting states, however, such property is usually difficult to achieve for most weak binding complexes. Several studies about encounter complexes based on chemical shift perturbation[175], nuclear spin relaxation[176], NOEs[177] and RDCs[178] have also been reported, besides, different types of NMR parameters can also be combined with some other biophysical approaches (such as SAXS[179]) for studying encounter complexes. Since most encounter complexes display very low binding affinities ($K_d > 100$ μM), NMR data analysis should be performed with care—it is usually impossible to obtain protein or ligand in $\sim 100\%$ bound form under normal sample conditions, therefore certain corrections are usually required to obtain data only corresponding to bound-form molecules[178].

1.3.3 Combining MD and NMR Approaches for Biomolecular Studies

MD and NMR are usually considered as two of the most important tools for molecular dynamics studies. Different NMR-based methods for studying protein dynamics have been reviewed in the previous section. Molecular dynamics (MD) is an *in silico* modeling approach to characterize dynamic behaviors of biomolecules with atomic-level details[180], and it can also provide certain information which is difficult to obtain experimentally. The earliest MD simulation study on biomolecules only reached ~ps timescale[181], however, it is already possible to record ~ms timescale MD trajectories for biomolecules nowadays with the aid of better computer hardwares and MD simulation programs[182]. It has been shown that certain information can be extracted

only from relatively long MD trajectories[183]. In recent years there has been much development for MD simulation on GPUs, which allows recording ~μs timescale MD trajectories in a routine fashion[184,185].

Since MD and NMR investigate from two different perspectives that are complementary to each other, there have been many efforts to combine these two approaches for studying biomolecules[186,187]. Different types of NMR parameters such as chemical shifts[188], RDCs[189], PREs[190] and PCSs[24] are commonly used as restraints in MD simulation programs for protein structure refinement. On the other hand, the accuracy of MD simulations largely depends on the quality of the underlying physical model (i.e., MD force field), which is in turn determined by the quality of input data for MD force field parameterization, therefore the availability of a variety of NMR data helps to develop better versions of MD force fields[191,192]. The quality of each MD force field is usually judged based on the agreement between MD simulation results and experimental data[193,194], different types of NMR parameters such as chemical shifts[195], scalar couplings[194] and order parameters[196] have been utilized for such purposes.

The combined MD and NMR studies can provide detailed pictures about the molecules being investigated. The microscopic dynamic behaviors of biomolecules can be directly visualized from MD trajectories, therefore MD simulation provides as a natural strategy for NMR data interpretation. On the other hand, the quality of MD force field can also get further improved from joint MD and NMR data analyses.

1.3.4 Summary of Original Contributions

In this work we have studied dynamic properties of two protein–peptide complexes based on Crk–Sos model system with combined MD and NMR approaches. These two complexes correspond to tight-binding and encounter complexes respectively. Those results about Crk–Sos encounter complex are more interesting since very few studies have been reported about encounter complexes formed by disordered proteins or peptides. Crk is an SH3-domain protein with $\sim$ 60 residues, Sos is a 10-residue peptide and it can be considered as a minimal model of disordered proteins. Certain properties

about this system can be learned from several previous studies: the X-ray structure of Crk–Sos tight-binding complex has been solved (PDB 1CKB[197]), the binding process of this complex has been studied via MD simulation based on a shorter version of Sos peptide[198]. The complex formed by Sos and wild-type c-Crk SH3 (wtSH3) displays relatively high binding affinity and thus belongs to the category of tight-binding complex. Some previous studies indicate that encounter complexes can be created by disrupting certain interactions that contribute to tight binding[199], such strategy has been followed to create modified Crk–Sos system for studying encounter complexes. Our MD and NMR results indicate that the re-engineered Crk–Sos complex displays many features similar to those observed on many other well-studied encounter complexes, besides, we have also identified potential targets for Amber ff99SB*-ILDN force field optimizations through joint MD and NMR data analyses.

The first part of this study focuses on the complex formed by wild-type c-Crk SH3 (wtSH3) and Sos. This complex shows relatively high binding affinity as indicated by previous ITC data ($K_d \sim 5\ \mu M$)[200], and it is also confirmed by our results measured on samples with lower ionic strength ($K_d \sim 1\ \mu M$). The large association rate constant k_{on} derived based on NMR titration data ($\sim 10^9\ M^{-1}s^{-1}$) indicates that strong electrostatic interaction is the major driving force for forming this complex. The ps~ns dynamics of Sos peptide within Crk–Sos complex is the central focus in this study, which is derived mainly based on complete sets of ^{15}N relaxation parameters (R_1, R_2, NOE and η_{xy}) measured under multiple magnetic field strengths. At the same time, we also carried out MD simulations for this complex with Amber ff99SB*-ILDN force field, which shows good agreement between MD simulation and NMR results in many previous studies[201–203]. However, we found that the agreement between MD simulation and NMR ^{15}N relaxation data was not perfect for this protein–peptide complex, and the largest discrepancy came from C-terminal arginines of Sos peptide: certain parameters from MD simulations (such as R_2 and NOE) are much higher than experimental data for those arginines, which strongly indicates that Sos arginines are involved in overstabilized salt bridges in MD simulations. We have tested several

alternative MD simulation setups such as using different water models or MD force fields, and we finally chose to modify specific Lennard–Jones (LJ) parameters in Amber ff99SB*-ILDN force field to solve this problem. Our modification focuses on fine-tuning LJ parameters of N^{η}–$O^{\delta}/O^{\varepsilon}$ atom pairs within Arg–Asp/Glu residue pairs. Specifically, the equilibrium distance σ between those atoms have been increased (specified by parameter λ) in order to weaken salt bridges formed on arginines. Such strategy already becomes effective as σ is increased by 3% (i.e., $\lambda = 1.03$), which causes very little perturbation to the original MD force field. Such problem has also been observed in some other studies based on small molecule systems recently[204,205], besides, the importance of LJ parameters in MD force field parameterization already starts to gain more attentions[206,207], therefore our MD force field modification strategy should point towards the correct direction for Amber ff99SB*-ILDN force field optimization in future[208].

The second part of this study focuses on the complex formed by double-mutant c-Crk SH3 (dmSH3) with Sos, which serves as a model system for studying protein–peptide electrostatic encounter complexes. This system is created based on the original Crk–Sos complex: two hydrophobic residues near the binding interface of Crk have been mutated (i.e., Y186L and W169F) with the aid of program FoldX[209]. The binding affinity of this protein–peptide complex is $K_d \sim 200$ μM as determined from NMR titration data, therefore this system is much better suited for studying electrostatic encounter complexes. We have carried out the same series of NMR measurements and MD simulations, most studies still focus on ^{15}N relaxation parameters of Sos peptide. However, the original Amber ff99SB*-ILDN force field still failed to produce MD simulation results in good agreement with NMR data: certain simulated ^{15}N relaxation parameters such as R_2 and NOE are much elevated over experimental data, which is still mainly caused by overstabilized salt bridges on Sos arginines. Our modified Amber force field ($\lambda = 1.03$) produces much better agreement for this re-engineered complex, therefore it provides further support to our MD force field modification strategy. In this protein–peptide encounter complex, Sos peptide be-

haves highly dynamic and it "dances" around multiple surface areas of dmSH3, such motional properties create a heterogeneous conformational ensemble of this complex. The joint MD and NMR analyses allow us to obtain more detailed dynamic pictures about protein–peptide encounter complexes based on Crk–Sos model system, which cannot be easily obtained from some other types of structural models (such as those derived based on PREs). Since it already becomes quite routine to perform ~μs timescale MD simulations nowadays, such strategy of combining MD simulation and NMR ^{15}N relaxation data can also be used to probe structural and dynamic properties of many other encounter complexes involving disordered proteins or peptides.

CHAPTER 2. DOMAIN COOPERATIVITY OF MULTIDOMAIN PROTEINS INVESTIGATED THROUGH RESIDUAL DIPOLAR COUPLINGS (RDCS)

2.1 Introduction

Modular design is a hallmark of many eukaryotic proteins. A great number of eukaryotic proteins contain multiple domains, one typical example comes from those protein families that are involved in signaling transduction pathways, which normally contain several adaptor domains (such as SH2, SH3) that are connected through stretches of predominantly flexible linkers. Each adaptor domain within the multidomain protein context usually has the property that can bind with multiple interaction partners, therefore such multidomain architecture plays important roles in regulating many types of biological interaction networks such as signaling transduction pathways. The binding of each adaptor domain to its targets is typically characterized by relatively low binding affinity, therefore those protein complexes usually have large dissociation rates which ensure fast dynamic response to external stimuli. Due to such property that many domains can interact with multiple binding targets, this mechanism gives rise to sophisticated, non-linear signaling networks, efficiently serving the needs of higher organisms[3].

The adaptor protein c-Crk serves as a good example to illustrate such binding properties. In mammalian cells, c-Crk exists in two different isoforms due to alternative splicing. The smaller species, c-Crk I, are comprised of an SH2 domain followed by a $\sim$ 15-residue linker and an SH3 domain (N-SH3). The larger form, c-Crk II, additionally contains a $\sim$ 50-residue linker and an extra SH3 domain (C-SH3)[210]. Crk can bind to a variety of phosphotyrosine-containing proteins through its SH2 domain, most prominently p130Cas. A number of important signaling pathways converge at the p130Cas/Crk assembly, which contains multiple copies of Crk and thus has an

ability to bring together a number of different Crk binding partners. The important biological roles of such interconnected pathways in cell growth, motility, proliferation, adhesion, and apoptosis have been well documented[211].

From the structural perspective, one of the key questions for Crk and many other modular proteins is about the interaction between multiple domains. On one hand, when two adjacent domains are connected through a long and flexible linker, the two domain should behave as essentially two independent entities, therefore domain–domain coupling or domain cooperativity should be minimal. On the other hand, two domains can be significantly coupled when they are connected through relatively short or conformationally rigid linker. The properties about Crk already become clear: c-Crk I conforms more closely to the former scenario, while c-Crk II represents the second type of situation[210]. More specifically, two constitutive domains SH2 and N-SH3 within c-Crk I display significant motional freedom, whereas in c-Crk II the three domains, SH2, N-SH3, and C-SH3, form a compact structure centered around the piece of the inter-SH3 linker. This result highlights the role of the extra domain: the binding surface in Crk C-SH3 has an unusual amino-acid composition, which degrades its ability to bind PxxP ligands; at the same time, C-SH3 occludes the binding site in N-SH3, thus downregulating a number of signaling events[210,212,213]. Another domain rearrangement, which has the effect of shutting down the Crk-mediated signals, occurs upon phosphorylation of the tyrosine residue within the inter-SH3 linker. The subsequent intramolecular binding of the pYxxP motif to the SH2 domain abrogates most of the Crk interactions[210,214,215].

Aside from c-Crk, many other examples can also illustrate the relationship between domain cooperativity and biological functions of multidomain proteins. Calmodulin (CaM) contains two domains connected by the linker region. The linker region is relatively rigid in the absence of Ca^{2+}, but becomes much flexible when Ca^{2+} binds to this protein. Therefore calmodulin exists in two different conformational states characterized by a different degree of domain–domain coupling as determined by the linker region. These two states are associated with different biological functions[8].

Protein tyrosine kinase Syk has tandem SH2 (tSH2) domain architecture. The two SH2 domains are connected by the 45-residue linker region which is relatively rigid in the unphosphorylated form; however, it becomes much more flexible once the linker region residue Y130 is phosphorylated. This event directly causes the activation of Syk kinase, leading to many further downstream effects[9].

Although these examples all underscore the importance of domain–domain coupling, however, it is not easy to investigate domain interactions by straightforward structural methods, one of the main reasons is due to the fact that these interactions are often dynamic in nature, corresponding to diverse and short-lived conformational species which are difficult to characterize by most structural techniques such as X-ray crystallography. However, it turns out that NMR is one of the best approaches for studying multidomain proteins, and two different NMR-based strategies are most suitable to address this situation. The first method relies on backbone ^{15}N relaxation—domain cooperativity usually leads to the increase of rotational tumbling motion correlation time τ_R, which can be sensitively detected by relaxation-based experiments[9,216]. This effect has been explored in-depth for two-domain proteins constructed based on GB1 that are similar to those proteins employed in this work[22]. An accurate interpretation, however, is still challenging because spin relaxation is sensitive to many motional modes, including domain–domain motions[217,218], besides, another difficulty comes from the fact that many existing algorithms for τ_R prediction have limited accuracy[219]. The second method makes use of residual dipolar couplings (RDCs). One of the popular strategies for RDC-based study is to load one of the domains with paramagnetic ions to achieve weak alignment in external magnetic field. If the two domains are fully coupled, then the second domain without paramagnetic ions should display the same degree of alignment as the domain loaded with paramagnetic ions; conversely, if the two domains are uncoupled, then the second domain should show much different alignment properties from the first domain, usually with much smaller alignment magnitude[220,221]. While elegant and potentially useful, this

method faces certain technical difficulties since it depends on suitably engineered metal-binding sites or carefully designed chelate tags[222–224].

In this study we resort to the more conventional alignment methods that only involve steric alignment mechanism to investigate domain cooperativity based on RDCs, this type of alignment mechanism is more likely to provide good correlation between RDC parameters and domain cooperativity than other types of alignment mechanisms. We further focus on the important limiting case where a pair of non-interacting domains is connected through a variable-length flexible linker. To model such a system, we have constructed a series of chimera proteins consisting of two copies of chicken α-spectrin SH3 domain[225] (α-spc SH3) that are connected through a flexible $(GGSGG)_n$-type linker[55]. In the case of a very short linker, this construct represents a dumbbell-shaped molecule which should align efficiently in steric alignment media. Conversely, if the linker is infinitely long, then each SH3 domain should behave as an independent entity, besides, since each individual domain has nearly spherical shape, it should display minimum amount of steric alignment. Therefore it can be expected that the degree of alignment should change in a sigmoidal fashion: from high values (short linker) to low values (long linker). The broad goal of this work is to quantitate this dependence, with both *in silico* structural modeling and experimental methods. In this manner we seek to establish an important reference case—two-domain proteins with minimal domain-domain coupling (due to a trivial linker connectivity). Anything beyond that would point towards domain cooperativity, such as direct domain–domain interactions, an increasingly structured linker region or the combination of these two factors.

2.2 Theory

2.2.1 Basic Principles of Dipole–Dipole (DD) Interactions and RDCs

The dipole–dipole (DD) interaction originates from the through-space interaction between magnetic dipole moments. The classical dipole–dipole interaction energy E_{DD} between a pair of magnetic dipoles $\mu_{\mathbf{1}}$ and $\mu_{\mathbf{2}}$ is:

$$E_{DD} = \frac{\mu_0}{4\pi r^3}(\mu_{\mathbf{1}} \cdot \mu_{\mathbf{2}} - 3\mu_{\mathbf{1}} \cdot \frac{\mathbf{rr}}{r^2} \cdot \mu_{\mathbf{2}}) \tag{2.1}$$

$\mathbf{r}$ is the vector connecting the two magnetic dipoles, the $\mathbf{rr}$ tensor can be expressed as:

$$\mathbf{rr} = \begin{pmatrix} xx & xy & xz \\ yx & yy & yz \\ zx & zy & zz \end{pmatrix} = \begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix} \tag{2.2}$$

The quantum-mechanical counterpart of classical dipole–dipole interaction is the dipole–dipole interaction Hamiltonian $\hat{H}_{DD}$. In the most simple case, considering a pair of nuclear spins $I = 1/2$ and $S = 1/2$, the Hamiltonian $\hat{H}_{DD}$ can be expressed as the following form:

$$\hat{H}_{DD} = -\frac{\gamma_I \gamma_S \mu_0 \hbar}{4\pi^2 r_{IS}^3}\left(3\hat{\mathbf{I}} \cdot \frac{\mathbf{r}_{IS}\mathbf{r}_{IS}}{r_{IS}^2} \cdot \hat{\mathbf{S}} - \hat{\mathbf{I}} \cdot \hat{\mathbf{S}}\right) = -\frac{\gamma_I \gamma_S \mu_0 \hbar}{4\pi^2 r_{IS}^3}\hat{\mathbf{I}} \cdot \mathbf{D} \cdot \hat{\mathbf{S}} \tag{2.3}$$

γ_I and γ_S are gyromagnetic ratios of spins I and S, μ_0 is the permittivity of the free space, $\hat{\mathbf{I}}$ and $\hat{\mathbf{S}}$ are spin angular momentum operators of spins I and S, $\mathbf{r}_{IS}$ is the vector connecting spins I and S whereas r_{IS} is the distance, $\mathbf{D}$ is the dipolar tensor

which can be represented as 3×3 matrix (either in Cartesian coordinates (x, y, z) or spherical polar coordinates (r, θ, ϕ)):

$$\begin{aligned}
\mathbf{D} &= \begin{pmatrix} \frac{3x^2}{r_{IS}^2} - 1 & \frac{3xy}{r_{IS}^2} & \frac{3xz}{r_{IS}^2} \\ \frac{3xy}{r_{IS}^2} & \frac{3y^2}{r_{IS}^2} - 1 & \frac{3yz}{r_{IS}^2} \\ \frac{3xz}{r_{IS}^2} & \frac{3yz}{r_{IS}^2} & \frac{3z^2}{r_{IS}^2} - 1 \end{pmatrix} \\
&= \begin{pmatrix} 3\sin^2\theta\cos^2\phi - 1 & 3\sin^2\theta\cos\phi\sin\phi & 3\sin\theta\cos\theta\cos\phi \\ 3\sin^2\theta\cos\phi\sin\phi & 3\sin^2\theta\sin^2\phi - 1 & 3\sin\theta\cos\theta\sin\phi \\ 3\sin\theta\cos\theta\cos\phi & 3\sin\theta\cos\theta\sin\phi & 3\cos^2\theta - 1 \end{pmatrix}
\end{aligned} \tag{2.4}$$

Since most NMR studies for biomolecules are carried out under strong magnetic fields (usually $\geq$ 500 MHz), the high-field condition can always be satisfied. Under the high-field limit condition, only the secular part of $\hat{H}_{DD}$ that is proportional to $\hat{I}_z\hat{S}_z$ remains important[1], therefore the truncated form of Hamiltonian $\hat{H}_{DD}^{sec}$ can be considered instead:

$$\hat{H}_{DD}^{sec} = -\frac{\gamma_I\gamma_S\mu_0\hbar}{4\pi^2 r_{IS}^3}(3\cos^2\theta_{IS} - 1)\hat{I}_z\hat{S}_z = b_{IS}(3\cos^2\theta_{IS} - 1)\hat{I}_z\hat{S}_z \tag{2.5}$$

and the dipole–dipole coupling constant b_{IS} is defined as:

$$b_{IS} = -\frac{\gamma_I\gamma_S\mu_0\hbar}{4\pi^2 r_{IS}^3} \tag{2.6}$$

which already assumes that RDCs are measured for a pair of nuclei with fixed distance r_{IS}. This constant b_{IS} reflects the magnitude of dipole–dipole interactions and is usually on the order of $\sim$ 10 kHz, which is much larger than many other types of interactions related to NMR studies (such as J-couplings), therefore most other terms can be neglected as DD interactions dominate.

[1]Assume I and S are different spin species.

The effects of DD interactions can be explained with the aid of perturbation theory, which treats $\hat{H}_{DD}$ as an additional perturbation term to the Hamiltonian of the whole system, since the absolute magnitude of $\hat{H}_{DD}$ (on the order of $\sim$ 10 kHz) is much smaller than the magnitude of Zeeman interactions $\hat{H}_Z$ (on the order of $\sim$ 100 MHz) under high-field conditions. The results from first-order perturbation theory indicate that the effects of $\hat{H}_{DD}^{sec}$ is to shift energy levels of nuclear spin eigenstates (i.e., Zeeman states), such that it modifies the transition energies between different energy levels, which usually manifests as shifted peak positions in NMR spectra. There are also many other effects originating from DD interactions, for example, DD interaction is usually the dominating nuclear spin relaxation mechanism for most spin species under solution state conditions, and its effects can be quantitatively calculated based on the second-order perturbation theory (i.e., Redfield theory[34]).

According to Equation (2.5), the magnitude of DD interaction is mainly determined by θ_{IS}, which is the angle formed between $\mathbf{r}_{IS}$ and external magnetc field $\mathbf{B}_0$. For SSNMR study, DD interaction is usually the most important factor that affects the appearance of NMR spectra—each molecule within solid-state samples has fixed orientation and thus fixed θ_{IS}, therefore when considering the ensemble of molecules within each sample, the distribution of $\theta_{IS} \in [-\pi, \pi]$ leads to the variation of DD interaction strength as $[-b_{IS}, 2b_{IS}]$ according to Equation (2.5), which causes the commonly observed powder pattern spectra[33]. DD interactions have found many different applications in SSNMR study, for example, most coherence transfer steps in SSNMR pulse sequences are based on cross-polarization (CP)[226], which usually makes use of DD interactions since its magnitude is much larger than other types of interactions such as J-couplings. DD interactions can also be used for internuclear distances measurement, such as the REDOR experiment[227]. However, in most cases it leads to much unwanted effects, especially causing severe line broadening in NMR spectra. In order to overcome such problems originating from DD interactions, many different

strategies have been developed to average out $\hat{H}_{DD}^{sec}$, in that case the time-averaged Hamiltonian $\langle \hat{H}_{DD}^{sec} \rangle$ should be evaluated instead:

$$\langle \hat{H}_{DD}^{sec} \rangle = b_{IS} \langle (3\cos^2\theta_{IS} - 1)\hat{I}_z\hat{S}_z \rangle \tag{2.7}$$

According to Equation (2.7), the averaging can be carried out either on spin part (i.e., $\langle \hat{I}_z\hat{S}_z \rangle$) or spatial part (i.e., $\langle 3\cos^2\theta - 1 \rangle$). Most earlier experiments perform averaging on the spin part, such as WHH-4[228] pulse sequence, and the effects can be well understood with the aid of average Hamiltonian theory[229]. With the discovery of magic angle spinning (MAS)[230], the averaging scheme has shifted to focus on the spatial part, and the main idea is to perform fast spinning for solid state samples under the "magic angle" condition to achieve $\langle 3\cos^2\theta_{IS} - 1 \rangle = 0$, such that $\langle \hat{H}_{DD}^{sec} \rangle = 0$. The averaging on spin and spatial parts can also be carried out at the same time in order to achieve even better results.

In solution state condition, there already exists one natural mechanism to average out DD interactions, which originates from the fast molecular overall tumbling motion on ~ns timescale, and it averages out $\hat{H}_{DD}^{sec}$ mainly through the spatial part. For isotropic overall tumbling motion, it causes $\langle H_{DD}^{sec} \rangle = 0$ such that first-order effects (i.e., peak position shifts) cannot be observed. However, if the tumbling motion is not completely isotropic (such as in the presence of alignment media or paramagnetic ions), it would cause $\langle \hat{H}_{DD}^{sec} \rangle \neq 0$ such that residual amount of DD interactions can still be observed, which are termed as RDCs.

2.2.2 General Concept of Alignment Tensor for Interpreting RDCs

The effects of RDCs can be much better understood by introducing the concept of alignment tensor $\mathbf{A}$ (or Saupe matrix $\mathbf{S}$), which describes the molecular orientation anisotropy under solution conditions[32]. In order to define the alignment tensor, a coordinate frame that is fixed to the molecule should be introduced first. Suppose the angles formed by the molecular frame with respect to external magnetic field

$\mathbf{B}_0$ are (ξ_x, ξ_y, ξ_z), the angles formed by vector $\mathbf{r}_{IS}$ with respect to three axes of the molecular frame are $(\zeta_x, \zeta_y, \zeta_z)$, then following relationship can be utilized

$$\cos\theta_{IS} = \cos\zeta_x \cos\xi_x + \cos\zeta_y \cos\xi_y + \cos\zeta_z \cos\xi_z \tag{2.8}$$

such that $\langle 3\cos^2\theta - 1\rangle$ can be rewritten as the following form:

$$\left\langle 3\cos^2\theta_{IS} - 1\right\rangle = 3\left\langle (\cos\zeta_x \cos\xi_x + \cos\zeta_y \cos\xi_y + \cos\zeta_z \cos\xi_z)^2\right\rangle - 1 \tag{2.9}$$

Under the assumption that the molecule being studied is totally rigid, $(\zeta_x, \zeta_y, \zeta_z)$ do not change since the defined coordinate frame is always fixed with the molecule. However, (ξ_x, ξ_y, ξ_z) may change with time due to molecular tumbling motions, therefore the averaging only need to be carried out for those angles:

$$\begin{aligned}\left\langle 3\cos^2\theta_{IS} - 1\right\rangle &= 3[\langle\cos^2\xi_x\rangle\cos^2\zeta_x + \langle\cos^2\xi_y\rangle\cos^2\zeta_y + \langle\cos^2\xi_z\rangle\cos^2\zeta_z \\ &+ 2\langle\cos\xi_x\cos\xi_y\rangle\cos\zeta_x\cos\zeta_y + 2\langle\cos\xi_x\cos\xi_z\rangle\cos\zeta_x\cos\zeta_z \\ &+ 2\langle\cos\xi_y\cos\xi_z\rangle\cos\zeta_y\cos\zeta_z] - 1\end{aligned} \tag{2.10}$$

The above relationship can be much simplified with the aid of alignment tensor $\mathbf{A}$, which is defined as (δ_{ij} is Kronecker delta)[32]:

$$A_{ij} = \frac{3}{2}\langle\cos\xi_i\cos\xi_j\rangle - \frac{1}{2}\delta_{ij} \tag{2.11}$$

such that observable RDCs can be written as

$$D_{IS} = 2b_{IS}\sum_{i,j=\{x,y,z\}} A_{ij}\cos\zeta_i\cos\zeta_j \tag{2.12}$$

In general, all elements of $\mathbf{A}$ are non-zero for arbitrarily chosen molecular frame, however, according to the definition of Equation (2.11), since $\mathbf{A}$ is real symmetric matrix, there exists certain similarity transformation to transform $\mathbf{A}$ as a diagonalized

matrix, and such transformation simply corresponds to the Euler rotation $R(\alpha, \beta, \gamma)$ that rotates from the arbitrarily chosen molecular frame to a special frame called principal axis system (PAS). In PAS only the diagonal elements A_{xx}, A_{yy}, A_{zz} are non-zero, therefore the relationship between RDCs and alignment tensor $\mathbf{A}$ becomes much simpler (the orientation of $\mathbf{r}_{IS}$ is expressed in spherical polar frame as (θ, ϕ)):

$$D_{IS}(\theta, \phi) = 2b_{IS}[A_{zz}\cos^2\theta + A_{xx}\sin^2\theta\cos^2\phi + A_{yy}\sin^2\theta\sin^2\phi] \tag{2.13}$$

or

$$D_{IS}(\theta, \phi) = 2b_{IS}[A_a(3\cos^2\theta - 1) + \frac{3}{2}A_r\sin^2\theta\cos 2\phi] \tag{2.14}$$

by convention it always has $|A_{xx}| \leq |A_{yy}| \leq |A_{zz}|$. $A_a = A_{zz}/2$ is the axial component of alignment tensor and $A_r = (A_{xx} - A_{yy})/3$ is the rhombic component. $R = A_r/A_a$ is defined as the rhombicity.

Many other parameters can also be defined based on the alignment tensor $\mathbf{A}$. In this work two other important parameters that have been utilized are GDO and NSP, both are defined based on the concept of scalar product between two alignment tensors. Since $\mathbf{A}$ is 3×3 real symmetric matrix and traceless, it only has five independent elements and can be treated as five-dimensional vector. In general, the scalar product between any two alignement tensors $\mathbf{A}^1$ and $\mathbf{A}^2$ can be defined as:

$$\langle \mathbf{A}^1 | \mathbf{A}^2 \rangle = \sum_{i,j=\{x,y,z\}} A^1_{ij} A^2_{ij} \tag{2.15}$$

with this definition, the parameter Generalized Degree of Order (GDO)[231] that characterizes the overall magnitude of alignment for a given sample is defined as:

$$\text{GDO} = \sqrt{(2/3)\langle \mathbf{A}^1 | \mathbf{A}^1 \rangle} = |A_a|\sqrt{1 + (3/4)R^2} \tag{2.16}$$

and Normalized Scalar Product (NSP) that serves as a generalized measure for relative orientation between two alignment tensors is defined as:

$$\mathrm{NSP} = \frac{\langle \mathbf{A}^1 | \mathbf{A}^2 \rangle}{\sqrt{\langle \mathbf{A}^1 | \mathbf{A}^1 \rangle \langle \mathbf{A}^2 | \mathbf{A}^2 \rangle}} \tag{2.17}$$

When the values of the normalized scalar product (NSP) is close to 1.0, it indicates that the two alignment tensors differ only by a scaling factor and thus "linear", whereas NSP values close to 0.0 suggest that the two alignment frames are "orthogonal". The situation where NSP $= -1.0$ can be described as "antiparallel" alignment: the rhombicities and alignment frames are in this case identical, while the degrees of alignment A_a have opposite signs. Although NSP defined by Equation (2.17) depends also on rhombicity R, it is mainly a measure of alignment axes orientation.

Since the introduction of alignment tensor makes the interpretation of RDCs much easily, it is usually the first step to determine alignment tensor based on RDCs, then many other further analyses can be carried out. The alignment tensor $\mathbf{A}$ is usually derived by fitting RDCs to already solved molecular structures with algorithms such as singular value decomposition (SVD)[232], and it is usually assumed that molecular structure used for the fitting is totally rigid. One important parameter that characterizes the quality of RDC fitting is the quality factor Q:

$$Q = \frac{\left(\sum_{i=1}^{N} (D_i^{expt} - D_i^{calc})^2 / N \right)^{1/2}}{\left(\sum_{i=1}^{N} (D_i^{expt})^2 / N \right)^{1/2}} \tag{2.18}$$

D^{expt} and D^{calc} are experimental and back-calculated RDCs respectively, N is the number of RDCs. In general, the fitting has better quality as Q approaches to zero. In certain cases, however, it is not possible to perform such fitting when the molecular structure is unknown, in those cases certain types of statistical analysis without the need of molecular structural information are more useful[233]. For most RDC prediction programs such as PALES[59], it is usually the first step to predict alignment tensor $\mathbf{A}$

based on molecular properties (such as overall shape and charge distribution), then RDCs can be calculated based on **A** and molecular structural information[59].

The alignment magnitude as well as RDCs are mainly determined by the following several factors: (i) When steric alignment mechanism dominates, molecular shape is the most important factor that determines alignment magnitude—molecules with near spherical shape tend to produce small alignment, whereas highly elongated molecules usually display much larger alignment. (ii) When electrostatic alignment mechanism dominates, the charge distribution at molecular surface becomes another important factor, and molecules can display significant alignment due to large asymmetric charge distribution at molecular surface. (iii) The concentration of alignment media also directly affects the amount of anisotropy in solution environment, therefore affects the alignment magnitude. In practice, since it is difficult to accurately measure very small RDCs, while too large RDCs may cause severe line-broading effects in NMR spectra, the magnitude of RDCs is usually fine-tuned around $\sim 0.1\%$ of b_{IS} in most cases, however, certain novel strategies have been developed such that very large RDCs can still be measured[234].

2.2.3 Guidelines for Domain Cooperativity Study Based on RDCs

The main idea of characterizing domain cooperativity based on RDC data can be described in the following way. For multidomain proteins, if we assume each domain itself is near spherical and the alignment mechanism is steric only, then each single domain should produce alignment magnitude near zero. This should also hold true when it is connected to other domains with a long and flexible linker, since in that case each domain still largely behaves as one single entity without being affected by other domains.

Now let us consider a generalized two-domain protein, where a pair of globular, near-spherical domains are connected via a flexible structureless linker. In the case when the linker is short, i.e., consists of no more than several residues, the protein has a dumbbell shape. Furthermore, a short linker allows for little motional degrees

of freedom such that domain–domain motion is highly restricted. To a good approximation, the protein can be described as one single rigid body with highly elongated shape. Being placed in steric alignment media, this protein should efficiently align and produce a high GDO value, besides, the long axis of the alignment tensor should roughly coincide with the line connecting the centers of mass of the two domains. Generally speaking, the alignment PAS in such a two-domain construct has no relation to the PAS of each individual domain that is near-spherical. Therefore the NSP between the alignment tensors derived for two-domain protein and single-domain protein should deviate significantly from 1.0 (unless by coincidence).

Then let us consider the opposite situation, where the flexible linker is very long. In this case the two domains should move almost freely and no longer sense each other's presence. Their GDOs, therefore, should approach that of an individual domain. Given that the shape of each domain is nearly spherical, the GDO should drop to a small value that is similar to one single domain. Likewise, the PAS orientation should generally approach that found in the individual domain. Consequently, the scalar product between the two alignment tensors—the one from the domain which belongs to the tandem and the one from the isolated domain—should converge towards 1.0 as the length of flexible linker region increases to infinity.

However, it should also be noted that neither GDO nor NSP is expected to show a quantitative convergence towards the ideal single-domain limit, since the behavior of a tethered domain is similar, but not identical to that of an isolated domain. In particular, if the shape of the domain is very close to spherical then even a small alteration of the shape, such as that caused by attachment of the linker, can lead to a significant change in magnitude and especially orientation of the alignment tensor. Therefore, the convergence towards the single-domain limit should not be viewed as a strict rule but should rather be treated as a broad trend, indeed such predicted trend has been observed in both *in silico* structural modeling and experimental studies, which will be presented with more details in later sections.

2.3 Theoretical Prediction Results

2.3.1 Two-domain Protein Structural Model Generation

In the *in silico* structural modeling study, several two-domain protein models have been generated, which are based on several single domains with different domain sizes. Since α-spc SH3 is also used for experimental study, more details about tandem α-spc SH3 structural model generation will be presented.

For tandem-SH3 domain construction, each individual α-spectrin SH3 domain was modeled based on the crystallographic structure 1U06[61] (resolution 1.49 Å). Protons and several missing heavy atoms from the side chain of residue D48 were added to the structure and optimized using the facilities of CHARMM[235] and Xplor-NIH[236] with CHARMM22 force field. The resulting domain coordinates were used to build structures of tandem dSH3 and several other constructs used in this study (such as fSH3). The structure 1U06 contains five side chains that are modeled with two alternate conformations. In the main structural model all of them are placed in the conformation found in the chain A of the crystallographic structure. Several other models have also been tested, in particular the one where the ionizable E17 side chain is placed in conformation B. The energy of this alternative structure, as evaluated in implicit generalized Born solvent[237], was found to be within 6 kcal/mol from the all-A structure. The structure, labeled BAAAA, was therefore deemed suitable for the analyses. In addition, the NMR structure PDB 1AEY[238] was also used.

As an example of the SH3 tandem, considering the structural model for dSH3-*ml*-dSH3, which is the two domain protein with 10-residue linker region. The amino acid sequence for this construct is M1(E7...L61)D62-GGGGSGGGG-(E7...L61)D62. The bracketed portions of the sequence correspond to the structured portion of the domains, as represented by the crystallographic coordinates 1U06. The remaining parts, including the 10-residue linker and single-residue termini, are modeled as random coils, which are generated with program TraDES[57] (subroutine foldtraj, default input parameters). For instance, in order to add N-terminal residue M1, a random

dipeptide with the sequence ME is generated first. This peptide is then "glued" onto the structure of the SH3 domain by overlaying the E residue of the peptide with E7 residue in the SH3 domain (specifically, N, C^{α}, and C' atoms are superimposed in this particular order). Finally, all redundant atoms are deleted, which include all atoms from the peptide residue E with the exception of H^{N}. The equivalent procedure is used for the C-terminal residue D62. In this case, we begin with the random dipeptide LD, which is "glued" to the SH3 domain by overlaying the L residue of the peptide with L61 from the SH3 domain (C', C^{α}, and O atoms are superimposed). Subsequently, all atoms from the placeholder peptide residue L are deleted. The same prescriptions are applied to insert the ten-residue linker segment, DGGGGSGGGG.

The dSH3-*ml*-dSH3 model obtained in this fashion is subjected to a further round of MD refinement (1 ps at 100 K), and the major goal is to relieve certain amount of strains caused by the structure assembly procedure. All atoms within each domain have been assigned an artificially large mass, 10 kDa, in order to preserve the relative position of the domains during this stage. The refinement protocol is implemented using the torsional angle dynamics (TAD) option in Xplor-NIH (IVM module[239]); the internal coordinates of the domains (all dihedral angles in the residues from 7 to 61, except for the angle ϕ in residue E7) are fixed and the planarity of peptide planes is enforced. This algorithm is used to produce an ensemble of 10000 conformers for dSH3-*ml*-dSH3. However, many conformers within the ensemble generated with this algorithm display severe steric clashes due to very close distances between certain atoms, and this issue cannot be resolved by the MD refinement procedure.

In order to exclude those conformers with steric clashes and only keep those meaningful structures within the ensemble, we evaluate the energy[2] for each conformer within the ensemble and analyze the resulting energy distribution. The results for a portion of the energy distribution for dSH3-*ml*-dSH3 structural model that contains 10000 conformers within the ensemble at the beginning is shown in Figure 2.1. The energies are evaluated using Xplor-NIH/CHARMM22 in vacuo. The low-energy part

[2]The energy includes all bonded and non-bonded interactions as calculated in vacuo.

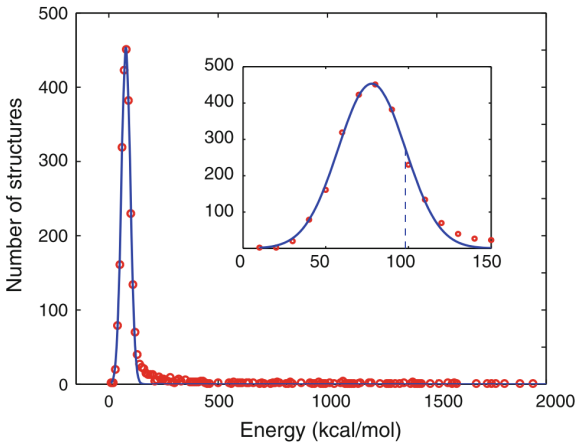


Figure 2.1. Part of the energy distribution for dSH3-*ml*-dSH3 model that contains 10000 conformers (red circles) fitted with the Gaussian function (blue curve). The inset displays the expanded portion of the main graph; the dashed line corresponds to one standard deviation σ of the Gaussian curve.

of the distribution follows the Gaussian curve (see inset in Figure 2.1). This is to be expected, since such energy distribution mainly arises from variability in the random-coil linker region. The Gaussian energy distribution for disordered protein systems can be predicted from general considerations and also has been derived from a number of theoretical models[240–242]. The results from random energy model (REM)[58] directly support the fact that Gaussian energy distribution should be observed on any kind of random polymers (such as disordered peptides).

In Figure 2.1, those structures that comply with the Gaussian distribution constitute a *bona fide* conformational ensemble, whereas the high-energy structures outside the Gaussian distribution (appearing to the right in the plot) suffer from steric conflicts. To address this problem, the energy distribution to the right of the dashed line (corresponding to one standard deviation σ of the fitted Gaussian curve) has been analyzed, and a large amount of structures that fall outside the curve were eliminated. In this manner the original ensemble of 10000 conformers was trimmed down to 2371 structures. The retained structures are (i) free of significant steric clashes and (ii)

have their two domains positioned according to the conformation of the (random-coil) linker, as generated by TraDES. Several randomly selected structures from the resulting reduced ensemble are shown in Figure 2.2.

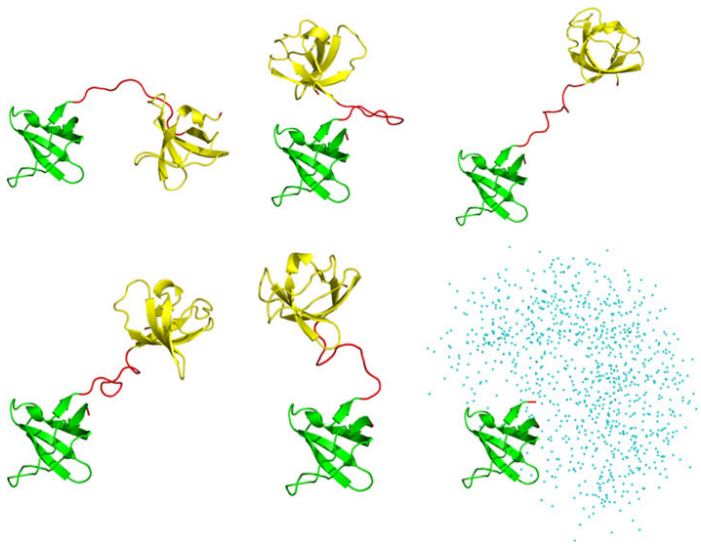


Figure 2.2. Five representative structures from the final dSH3-*ml*-dSH3 ensemble. The N- and C-terminal SH3 domains (backbone coordinates same as PDB 1U06, residues 7–61) are plotted in green and gold, respectively; the regions modeled as random coil are colored in red. The panel in the lower right corner shows the spatial distribution of the centers of mass of the C-terminal domain (blue dots) in the coordinate frame of the N-terminal domain (as plotted for the entire ensemble).

The same procedure as described above has also been applied to construct *in silico* structural models for several other chimera proteins, which include disintegrin domain (PDB 1MPZ[243], 41 residues, R_g = 8.8 Å), SH2 domain (PDB 1A81[244], 106 residues, R_g = 13.0 Å), and nucleotide-binding domain (NBD) (PDB 3GWI[245], 170 residues, R_g = 15.3 Å). Each domain selected for *in silico* model generation has near-sperical shape and thus displays minimum amount of steric alignment. We also generated *in silico* model for fSH3 (full-length α-spc SH3, 62 residues) and dSH3 (truncated α-spc SH3, 57 residues, residues 2–6 at N-terminal are deleted) in order to compare results

with two-domain protein constructs and calculate certain RDC parameters such as NSP.

2.3.2 RDC Predictions Based on *in silico* Structural Models

In this work PEG/hexanol alignment media have been used for all RDC data measurement, whose alignment mechanism is commonly assumed to be steric[38,246]. The program PALES was used to model this media, which was originally developed for DMPC/DHPC bicelles[36]. The thickness of the planar bilayer was set to 28.6 Å[247]. Since the bicelle calculations in PALES have been programmed for DMPC/DHPC, the code already implicitly accounts for the presence of 5 mM free DHPC in the solvent[248,249]. To correct for this small contribution and account for 0.3 wt% of free hexanol in solution[247], slightly altered value of the liquid crystal concentration should be used. For instance, in the PALES calculations aimed at 5% PEG, $r = 0.85$ media the effective liquid crystal concentration should be set to 65 mg/mL. The order parameter of the PEG/hexanol bicelle was assumed to be the same as for DMPC/DHPC, which is 0.8.

Several predictions of alignment parameters under the assumption that an electrostatic mechanism plays a significant role have also been performed. The corresponding PALES calculations involve several additional variables: charge density on the surface of the bilayer σ (treated as adjustable parameter), sample pH (actual value 3.5; also treated as adjustable parameter), and ionic strength of the solution (17 mM, corresponding to 20 mM citrate buffer at pH 3.5). The program PROPKA was used to determine the charges on α-spc SH3, which calculates the pK_a values of ionizable groups[250,251]. Since there is significant structural variability in tandem constructs (and also fSH3), PROPKA calculation was performed for each individual conformer within the structural ensemble. The coordinates of a given conformer plus all partial charges calculated for this particular conformer were then used together as input for PALES calculations. More sophisticated methods for calculating charges on ioniz-

able groups have not been tested, since such effort seems unwarranted given that the models used for prediction of electrostatic alignment remain relatively crude[252–254].

RDCs predicted for the individual conformers were averaged to produce one single set of simulated RDCs for each structural ensemble. When drawing comparison with the experimental data from the tandem constructs, the computed RDCs from the N- and C-terminal domains within two-domain protein construct were further averaged. Such averaging procedure is fully justified so long as both domains have the same and invariant internal structure, since by definition RDCs already represent the average between multiple molecular orientations. The resulting dataset is subsequently restricted to the core portion of the α-spc SH3 domain, residues 9–60, and fitted with the protonated structure 1U06. In the situation when residues 9–60 are represented by one and the same set of coordinates throughout the course of the calculations, the quality factor Q from such fitting procedure is always zero. The alignment parameters extracted in this fashion are used to calculate the theoretical values of GDO and NSP according to Equations (2.16) and (2.17).

2.3.3 Dependence of RDCs with Respect to Domain Cooperativity

The results of PALES predictions on several *in silico* structural models of two-domain proteins are shown in Figure 2.3. As expected, as the length of linker region increases, it leads to pronounced decreases in the GDO values. Eventually, each curve approaches a plateau which is close to, but distinct from the GDO of the isolated domain (dashed horizontal lines in Figure 2.3A). Similarly, as the linker length is increased, the NSP value becomes closer to 1.0 (Figure 2.3B), however, this trend is relatively weak, which means that the orientation of the alignment frame in the tethered domain remains quite different from the isolated domain.

It is interesting to note how the results scale with the size of the system since those single domains used for *in silico* structural modeling have quite different sizes. Generally, it is expected that for the smallest unit, disintegrin, the two domains become "uncoupled" and the GDO/NSP plateau is reached already with a moderately

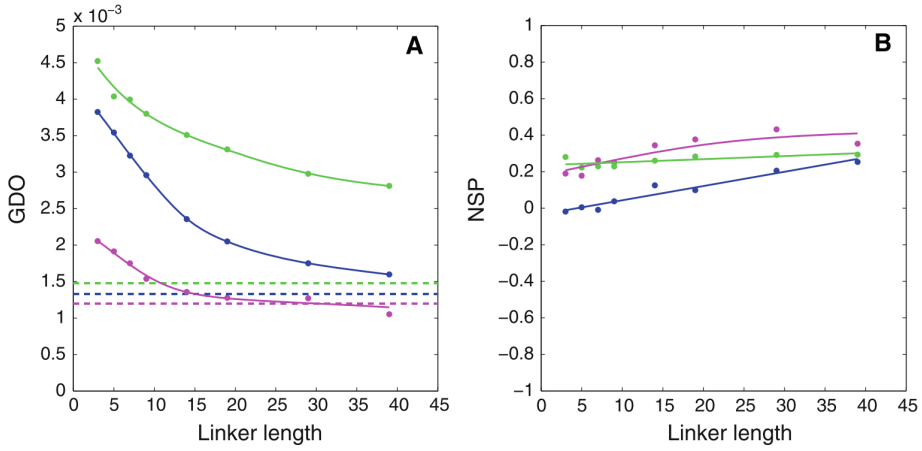


Figure 2.3. PALES simulations of steric alignment in a series of computer-generated tandem proteins (disintegrin: magenta; SH2: blue; NBD: green). (A) Simulated GDO values for domains which belong to the tandems (circles) and the corresponding isolated domains (dashed lines). (B) Simulated NSP values that characterize the mutual projection of the alignment tensors from domains within the tandem and the corresponding isolated domains.

long linker. Conversely, for the biggest unit, nucleotide-binding domain (NBD), this does not happen until the linker becomes much longer. A rough theoretical estimate can be readily obtained for the scaling properties of the two-domain constructs. The R_g for disintegrin and NBD domains are 8.8 Å and 15.3 Å respectively, for simplicity it can be assumed that both domains are spherical. The average end-to-end distance in the random peptide chain, such as the computer-generated linker, scales with the number of residues as $N^{0.6}$ [255]. From here the "scaling factor" can be estimated that relates the NBD tandem to disintegrin tandem. For example, the NBD tandem with 25-residue linker can be viewed as a "resized" copy of the disintegrin tandem with 10-residue linker. This is roughly consistent with the trends reflected in Figure 2.3A.

Generally speaking, N- and C-terminal domains in the tandem are inequivalent: one is attached to the linker through its C-terminal residues while the other through the N-terminal residues. In principle, this should lead to doubling of the spectral resonances. This situation is illustrated in Figure 2.4, which shows two domains

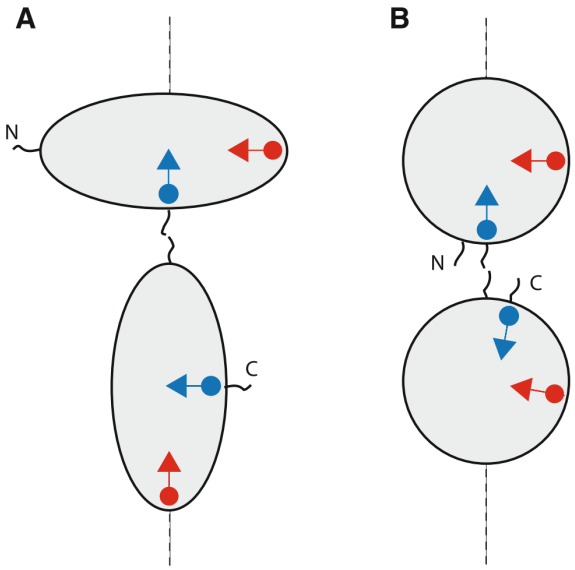


Figure 2.4. Schematic structure of tandem domains with resident spins. Blue circles represent ^{15}N spins belonging to i-th residue in the N- and C-terminal domains (assuming that residue numbering is the same in the two domains); blue arrows represent the corresponding ^{15}N–^{1}H^{N} vectors. Red circles/arrows refer to the j-th residue. The dashed line corresponds to the long axis of the steric alignment tensor.

connected in tandem. The scheme Figure 2.4A suggests that some of the amide signals from N- and C-domains should be resolved with respect to their chemical shifts. Furthermore, most of the signals should be distinguishable with respect to dipolar couplings. For instance, the equivalent residues from N- and C-terminal domains (blue circles) sense somewhat different chemical environment: one is located next to the fused terminus, while the other is close to the free terminus. Hence it can be expected that two distinct HSQC peaks should be observed from these sites. Moreover, the orientation of the ^{15}N–^{1}H^{N} vectors at the two sites (blue arrows) relative to the long axis of the alignment tensor (dashed line) is also different. Hence RDCs for these two sites should be different as well. In fact, it is anticipated that most of the N- and C-terminal residues can be differentiated on the basis of their dipolar

couplings. For instance, another pair of equivalent residues (red circles) illustrates the case where chemical shifts are likely to be degenerate, but RDCs are very different.

However, there may also exist some alternative cases, for example, Figure 2.4B represents a "softer" version of the same scenario. When N- and C-termini of each domain are positioned close to each other, RDCs from the two domains may turn out to be very similar. Apparently, this scenario holds true for the α-spc SH3 tandem according to our experimental data, besides, similar results have been reported in the experimental study of GB1 construct featuring a short 3-residue linker[22]. Among all of those *in silico* structural models have been constructed, it turns out that NBD tandem illustrates the situation where the couplings from the two domains are reasonably close (Figure 2.4B), while the other two constructs (based on disintegrin or SH2) predict significant differences between the two domains (Figure 2.4A).

2.4 Experimental Results

2.4.1 Sample Preparation

The original pET3d plasmid encoding chicken α-spectrin SH3 domain (i.e., full-length construct fSH3) was used as starting point to generate all other constructs. Since the N-terminus of α-spc SH3 is highly disordered[61,225,238], residues 2–6 at N-terminal were deleted to produce the deletion construct dSH3. Three chimera proteins have been constructed, each contains two copies of dSH3 connected via a glycine-serine linker. Three different linker regions have been used, with sequence as GSG, GG(GGSGG)GG, and GG$(GGSGG)_5$GG (the latter contains 29 residues). In what follows, these tandem constructs are referred to as dSH3-*sl*-dSH3, dSH3-*ml*-dSH3, and dSH3-*ll*-dSH3, indicating short, medium-length, and long linker, respectively. To produce the cDNA for tandem constructs, two separate PCR reactions were carried out. One PCR product encoded N-terminal dSH3 plus half of the linker region; it was digested with restriction enzymes *NcoI* and *BspEI*. The other PCR product encoded the remaining half of the linker plus C-terminal SH3 domain; it was digested with

BspEI and *BamHI*. In the next step, the two digested PCR products were cloned into pET3d vector and, at the same time, fused by means of the T4 DNA ligase. Note that the use of *BspEI* sites and T4 DNA ligase does not introduce any unwanted residues in the linker sequence. All plasmids were transformed into BL21(DE3) *E. coli* strain and sequenced to check for accuracy. More details about plasmid construction and primer design strategies are included in Appendix A.

The bacteria transformed with plasmids were grown in 1 L of M9 minimal media enriched with $^{15}NH_4Cl$ at the temperature 37 °C. Protein expression was induced at $OD_{600} \sim 0.6$ by adding 0.5 mM IPTG. After 4 h of incubation the cells were harvested by centrifugation (15 min at 8000 rpm) and stored at −80 °C. The cell pellet was then resuspended in pH 8.0 lysis buffer and lysed using a French Press. fSH3 was contained in the supernatant; it has been purified following the previously described protocol[61]. All other constructs formed inclusion bodies during the expression. To purify the precipitated protein material, the French press lysis pellet was first washed four times using 20 mM Tris (pH 8), 1 mM EDTA, 100 mM NaCl, 0.5% Triton X-100 buffer. Each wash was carried out in 30 mL of the buffer then centrifuge 30 min at 18000 rpm before discarding the supernatant; the last wash was done without the Triton. The pellet was then dissolved in 30 mL of 8 M urea, 20 mM Na-citrate, pH 3.5 denaturing buffer and heated in water bath at 70 °C for 30 min. The solution was subsequently dialyzed over the course of 1 day against 20 mM Na-citrate, pH 3.5 refolding buffer using the dialysis bag with MW cutoff 3 kDa. The refolded protein was purified using the Superdex 75 gel-filtration column and then concentrated using the Millipore Amicon devices with 3 kDa MW cutoff. All samples have been analyzed by SDS-PAGE to check the protein molecular weight. Protein concentration determined by UV absorbance at 280 nm was in the range from 0.5 to 1.0 mM for each sample. In addition to α-spectrin SH3 domain, two samples of human ubiquitin have also been prepared using expression and purification procedures adapted from previously described protocols[256].

2.4.2 RDC Measurement for Tandem α-spc SH3

The experimental measurements involved five constructs of α-spectrin SH3 domain (fSH3, dSH3, dSH3-*sl*-dSH3, dSH3-*ml*-dSH3, and dSH3-*ll*-dSH3) in solution with 20 mM Na-citrate, pH 3.5, 0.02% (w/v) NaN_3. One extra sample of fSH3 additionally included 100 mM NaCl (see Table 2.1). The oriented samples were prepared by using 5% pentaethylene glycol monododecyl ether (C12E5) and 1-hexanol[38]. Two series of samples were prepared with the C12E5/hexanol molar ratio $r = 0.85$ and 0.96, respectively. The former condition offers stable alignment around the working temperature 25 °C[38]; the latter is the most popular choice in the studies employing PEG/hexanol media[257–259]. Special care was taken to maintain the same alignment conditions throughout each series of samples. Towards this goal, large volume of stock solutions containing 10% C12E5 and $r = 0.85$ (or $r = 0.96$) fraction of 1-hexanol was prepared in advance. The anisotropic samples were prepared by mixing 150 μL aliquot of the stock solution with 150 μL aliquot of the protein solution to achieve 300 μL volume to be measured in Shigemi tube. Each freshly prepared batch of protein was used to make two or three NMR samples, including some replicate samples. Each sample was allowed to equilibrate in the magnet for 2 h, before the residual quadrupolar coupling $\Delta(^2H)$ from 10% D_2O/90% H_2O solvent was carefully measured. The $^1H^N$–^{15}N RDC data were collected using the IPAP-HSQC pulse sequence[60,260] on a Varian Inova 600 MHz spectrometer equipped with a triple-resonance probe and z-axis gradient. Data sets for isotropic and partially aligned samples were collected as (192, 576) complex matrices with spectral widths of 1920 and 9000 Hz in nitrogen and proton dimensions, respectively. Each experiment took approximately 2 h. After collecting each dataset $\Delta(^2H)$ was remeasured, and the drift never exceeded 0.2 Hz. The IPAP spectra were processed using the NMRPipe software package and peak positions were determined by means of the nlinLS fitting subroutine[261]. In the case of tandem constructs, most of the peaks from the N-terminal domain are neatly overlapped with their counterpart peaks from the C-terminal domain. Hence, the RDC values measured for these residues represent the average of the N- and C-

domain couplings. For a small number of sites that give rise to two resolved peaks, the experimental RDCs were averaged on a pairwise basis. In analyzing the results, all datasets have been restricted to the core portion of the SH3 domain, residues 9–60.

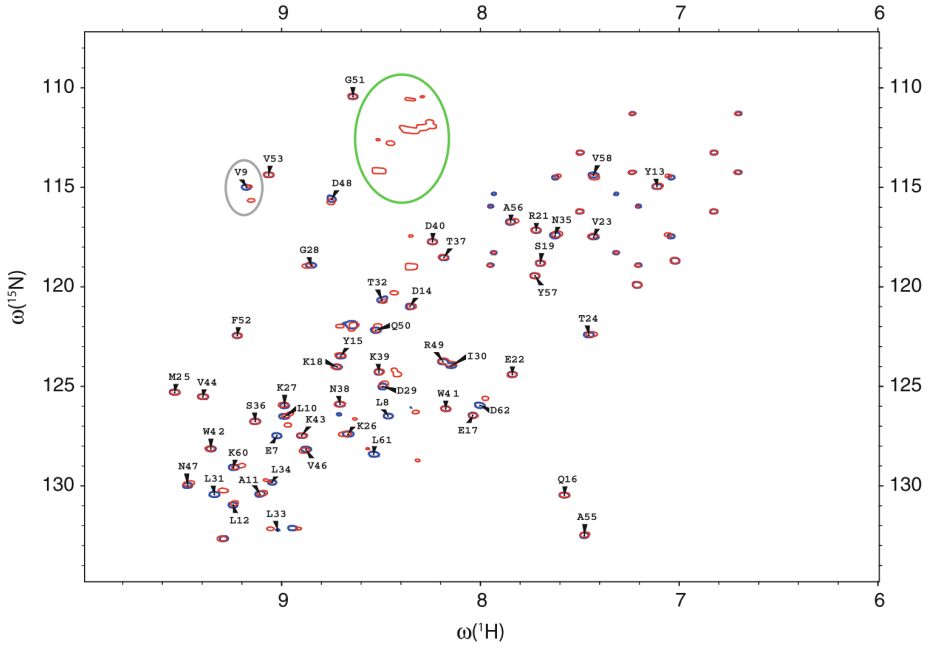


Figure 2.5. HSQC spectra of the isolated dSH3 domain (blue) and dSH3-*ll*-dSH3 tandem (red). The dSH3 resonances are labeled except for Trp, Asn, and Gln side-chain correlations. The elliptical contours outline the resonances from the glycine-rich linker and from the residue V9 which demonstrates peak doubling effect.

The spectra of the dSH3 tandems turn out to be very similar to the spectrum of the isolated dSH3. For example, Figure 2.5 shows the spectrum of the two-domain construct with long disordered linker dSH3-*ll*-dSH3 superimposed on the spectrum of the single domain dSH3 (spectral peaks colored red and blue, respectively). The resonances from multiple glycine residues of the random-coil $GG(GGSGG)_5GG$ sequence are all grouped in the area of the spectrum where they do not interfere with intra-domain peaks (outlined by the green contour). Most of the peaks from N- and

C-terminal domains overlap. Furthermore, they overlap or fall very close to the peaks from the isolated dSH3 domain (cf. red and blue contours in Figure 2.5), so that the assignments[238] can be transferred in a straightforward fashion. The lack of resonance shifts between the isolated domain and the tandem confirms the absence of specific domain–domain interactions.

According to results shown in Figure 2.5, only a small number of residues that are located close to the linker attachment point give rise to a pair of resolved peaks, corresponding to the sites in the N- and C-terminal domains. One typical example of this behavior is residue V9, as outlined by the grey contour in the spectral map Figure 2.5. The effect is most pronounced in the short-linker construct, dSH3-*sl*-dSH3, where the two domains are spatially close. In this sample, 18 residues show the peak doubling effect, with half of them fully resolved. In particular, the peaks from terminal residues 7, 8 and 61, 62 are strongly affected. Inspection of the crystallographic structure shows that these two segments are connected: there is a backbone–backbone hydrogen bond between residues 8 and 61. Thus all of the four residues prove to be in the immediate vicinity of the linker region, resembling the arrangement shown in Figure 2.4B. The split peaks associated with residues 7, 8, and 61, 62 show large chemical-shift separations, which makes it impossible to assign most of them. Furthermore, it is likely that the conformation of these residues in dSH3-*sl*-dSH3 differs from that seen in the crystallographic structure of fSH3. Considering these complications we choose to restrict the data analysis to the core region of the domains, i.e., residues 9–60. This restriction has been applied to all α-spc SH3 constructs investigated experimentally as well as those *in silico* structural models.

The results as shown in Figure 2.5 can also partially answer the question that whether resonances from N- and C-terminal domains can be distinguished based on RDCs or not. In principle, it can be expected that a pair of peaks that is completely overlapped in HSQC spectrum would become resolved in IPAP-HSQC (due to potentially large differences in RDCs, on the order of tens of Hz; see Figure 2.4A). In reality, however, such behavior has not been observed despite the high resolution of

the IPAP-HSQC spectra. In the experiments involving dSH3-*sl*-dSH3, a few RDCs that have been measured separately for the sites in the N- and C-domains differ by no more than several Hz (e.g., in the case of V9 the difference is 0.5 Hz). At this level, RDCs alone cannot produce resolved spectral peaks. In summary, it appears that all tandem dSH3 constructs behave as indicated in the scheme of Figure 2.4B: the N- and C-termini of each domain are aligned and together form a linker attachment point, therefore RDCs in the two domains are similar.

Given that only few residues allow for separate measurements of RDCs associated with N- and C-terminal domains and the measured values prove to be very similar, such couplings have all been averaged and then later added to form the bigger dataset consisting of all average RDCs. For instance, in the case of dSH3-*ll*-dSH3 there are 44 couplings measured that inherently represent the average values, plus two distinct couplings from the residue V9. After these two couplings are averaged, the final data set consists of 45 entries (pertaining to the region 9–60 in the dSH3 domain). The same approach has been applied for all other tandem-SH3 samples.

2.4.3 Alignment Tensor Derivation Based on RDCs

The RDC data obtained for all constructs based on α-spc SH3 have been fitted to the crystallographic coordinates 1U06. Although the experimental RDCs from the tandem constructs represent the averages between the N- and C-terminal domains, they have been fitted in the standard fashion, the same as those data from the isolated dSH3 and fSH3 domains. As already commented, this approach is fully legitimate so long as the internal structure of the two domains is identical. The quality of the RDC fitting for dSH3 and dSH3-*ll*-dSH3 is illustrated in Figure 2.6A,B. In both cases the agreement between the experimental and fitted couplings is reasonably good, as evidenced by the quality factor Q of 0.25 and 0.26 (typical for a small globular protein represented by a crystallographic structure with 1.5 Å resolution[262]).

Two residues consistently show poor agreement in all fits—N47 and D48 (grey circles in Figure 2.6). This discrepancy should be attributed to the local conformational

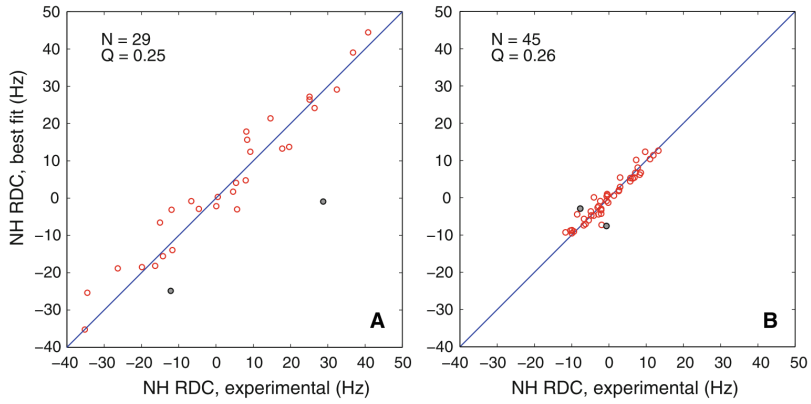


Figure 2.6. Fits of the experimental RDCs from (A) the isolated dSH3 domain and (B) dSH3-*ll*-dSH3 tandem with the structural model based on the coordinate set 1U06 (protonated, energy minimized). The fitting is limited to the core portion of the protein (residues 9–60) and does not include the data from N47 and D48 (grey circles in the plot). The total number of the fitted RDCs and the quality factor Q are indicated in the graphs.

variability at this particular site. In the crystal, N47 assumes an unusual conformation ($\phi \approx 50°$, $\psi \approx -110°$)[263] which has a significant destabilizing effect on the structure[264]. A similarly unfavorable conformation is seen in solution structure ($\phi \approx 60°$, $\psi \approx -75°$), except for one conformer[238]. The built-in strain apparently triggers local motions at this site. Both N47 and D48 display sharply elevated B-factors and several of the side-chain atoms in D48 cannot be resolved by X-ray crystallography at all. D48 is also the only residue featuring exchange broadening in solution ($R_{ex} = 28\ s^{-1}$ at 600 MHz)[265]. As it turns out, the RDC data from N47 and D48 can be accommodated using the existing structural models. Specifically, if these two residues are "grafted" from the NMR structure 1AEY onto the crystallographic structure 1U06, some of the resulting hybrid models lead to a satisfactory RDC fit.

The alignment parameters derived from multiple SH3 samples are summarized in Tables 2.1, 2.2 and in Figure 2.7 (the complete RDC datasets are shown in Tables A.1, A.2). As has already been mentioned earlier, special care was taken to ensure that the alignment conditions are the same throughout each series of samples. In the

Table 2.1.
Alignment parameters from a series of samples in $r = 0.85$ PEG/hexanol media

	dSH3[b]	dSH3[a]	fSH3	fSH3 (high salt)[c]	dSH3-*sl*-dSH3	dSH3-*sl*-dSH3[a]	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3[b]	dSH3-*ll*-dSH3[a]
A_a (10^{-3})	2.23	3.01	−1.40	−2.27	−1.56	−1.51	0.67	0.65	0.75
R	0.62	0.40	0.37	0.39	0.37	0.36	0.65	0.27	0.29
$\{\alpha, \beta, \gamma\}$[d]	308° 62° 81°	308° 65° 80°	32° 38° 320°	27° 40° 325°	33° 80° 63°	34° 80° 63°	168° 53° 159°	107° 54° 155°	106° 54° 155°
N	29	25	46	27	42	42	39	45	45
Q	0.25	0.30	0.22	0.21	0.22	0.21	0.27	0.26	0.27
$\Delta(^2\text{H})$(Hz)	24.9	21.8	24.7	25.8	23.3	23.1	16.9	23.5	24.4
GDO (10^{-3})[e]	2.03[f]	2.93	1.19[f]	1.85	1.41[f]	1.37	0.92[f]	0.57[f]	0.63
NSP[g]	1.00	0.98	0.94	0.94	−0.53	−0.53	−0.32	0.26	0.27

[a] Duplicate samples

[b] The RDC fits are shown in Figure 2.6

[c] 100 mM NaCl added to the buffer

[d] Relative to the crystallographic coordinates 1U06

[e] Normalized towards $\Delta(^2\text{H}) = 20$ Hz

[f] Values used to optimize the electrostatic alignment model

[g] Relative to dSH3

case of $r = 0.85$ samples, the residual quadrupolar splittings $\Delta(^2\text{H})$ proved to be fairly uniform and reproducible (with the apparent exception of dSH3-*ml*-dSH3, see Table 2.1). For $r = 0.96$, the splittings were lower than expected and varied significantly from one sample to the other (see Table 2.2). This latter media appears to be only marginally stable at 25 °C[38]. Generally, the dilute PEG/hexanol media used in NMR

Table 2.2.
Alignment parameters from a series of samples in $r = 0.96$ PEG/hexanol media[a]

	dSH3	fSH3	dSH3-*sl*-dSH3	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3
A_a (10^{-3})	2.23	−0.78	−1.50	−0.78	0.41
R	0.40	0.33	0.36	0.64	0.41
$\{\alpha, \beta, \gamma\}$	306° 65° 79°	32° 39° 321°	33° 80° 63°	36° 83° 64°	103° 54° 154°
N	43	48	41	40	45
Q	0.25	0.21	0.22	0.26	0.25
$\Delta(^2\mathrm{H})$(Hz)	17.1	12.3	20.3	15.0	11.0
GDO (10^{-3})	2.76	1.32	1.54	1.18	0.80
NSP	1.00	0.90	−0.64	−0.47	0.21

[a] The conventions are the same as in Table 2.1

experiments falls close to the boundary between lamellar phase L_α and a "sponge phase" L_3 , which itself has a complex and variable morphology[266]. In this situation, minor variations in the composition of the sample—possibly, the differences between the protein constructs themselves—may cause significant changes in the alignment. To account for these changes, we choose to normalize the GDO parameter according to the magnitude of $\Delta(^2\mathrm{H})$, see Table 2.1. In doing so we assume that $\Delta(^2\mathrm{H})$ reflects the aligning properties of the media not only with respect to water, but also to different α-spectrin SH3 constructs. This is an admittedly ad hoc approach, with only limited experimental justification[38]. However, from another perspective such strategy should still be valid for studying a series of samples where the alignment is dictated by the invariant properties of the α-spc SH3 domain. At least in the case of $r = 0.85$ media the alignment conditions seem to be highly reproducible from one sample to another.

The consistency of different sample preparations can be verified through measurements on replicate samples. The results from the duplicate dSH3-*sl*-dSH3 and

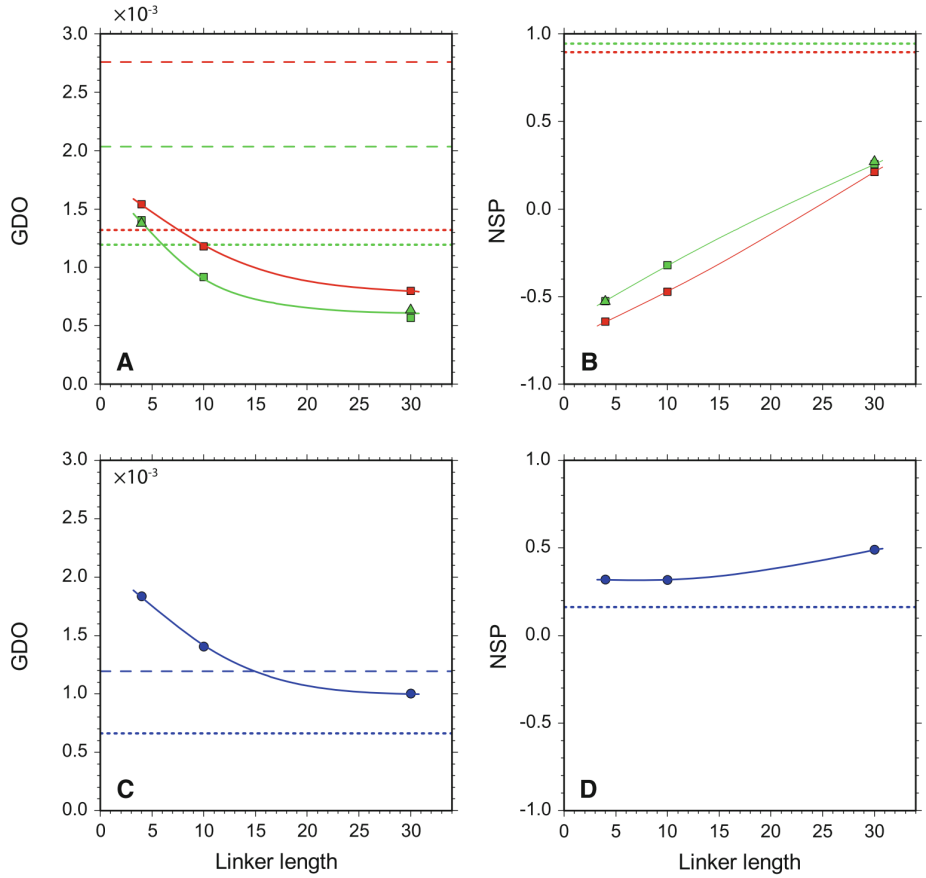


Figure 2.7. GDO and NSP alignment parameters in a series of constructs of α-spc SH3 domain. (A, B) Experimental data from dSH3-*sl*-dSH3, dSH3-*ml*-dSH3, and dSH3-*ll*-dSH3 tandem constructs (squares for the main series of samples, triangles for replicate samples), as well as dSH3 and fSH3 single-domain constructs (dashed and dotted horizontal lines, respectively). The data from r = 0.85 and 0.96 samples are indicated in green and red, respectively. The NSP parameter is calculated relative to the dSH3 construct (the corresponding value for the dSH3 itself is 1.0). (C, D) The corresponding PALES-based simulations using the steric alignment model which assumes planar bicelle bilayers with the thickness 28.6 Å.

dSH3-*ll*-dSH3 samples are in very good agreement with each other (see Table 2.1; also compare squares and rectangles in Figure 2.7). Furthermore, the data taken at r = 0.85 and 0.96 are also in good agreement insofar as rhombicity and the orientation of the alignment frame are concerned. One apparent exception is dSH3-*ml*-dSH3,

cf. Tables 2.1 and 2.2. This construct displays the rhombicity R approaching the maximum value of 2/3. Under these circumstances, the labeling of the principal axes of the alignment tensor becomes ambiguous, causing apparent differences in the Euler angles. In fact, the two data sets are highly consistent, as confirmed by the NSP of the respective alignment tensors, 0.998.

The only sample where reproducibility truly appears to be an issue is dSH3. While the orientation of the alignment frame is well reproduced, the rhombicity shows a significant amount of variation as shown in Tables 2.1 and 2.2. Furthermore, the GDO value varies significantly from one preparation to the other. Besides, one important feature can be recognized is that the alignment of dSH3 is unexpectedly strong in all anisotropic samples, which will be discussed in later sections with more details.

Figure 2.7A provides the experimental verification of the concept presented in this work: the tandem with a short linker should align strongly, whereas the tandem with a long linker should align more weakly (given that the shape of the domains is close to spherical, the linker is unstructured, and there is no significant domain–domain interactions). Furthermore, it appears that the simple steric alignment model, as implemented in PALES, in conjunction with the conformational ensembles generated with the help of TraDES, provide an adequate explanation for the experimental observations (compare squares/triangles in Figure 2.7A and circles in Figure 2.7C). Note that the PALES-based prediction algorithm does not involve any adjustable parameters. The NSP data in Figure 2.7B confirm our hypothesis that with increase in linker length the tandem domains begin to behave as independent entities, i.e., align similarly to isolated domains.

However, the results in Figure 2.7 also indicate that the steric alignment model is only partially successful in reproducing the experimental data. While the GDO values for tandem constructs are predicted with reasonable accuracy, the predictions for isolated domains, dSH3 and fSH3, are off the mark (cf. dashed/dotted lines in Figure 2.7A,C). The case of the dSH3 is particularly instructive. This domain represents a spheroid with modest anisotropy, $I_{\parallel}/I_{\perp} = 0.83$. Its structure is well defined: only one

N-terminal and one C-terminal residue are classified as "unstructured". Assuming that the mechanism of alignment is steric, dSH3 should show only a modest degree of alignment, yet the experimental GDO value is exceedingly high. Moreover, the experimentally determined alignment frame bears no similarity to the predicted one, as indicated by the low NSP value, −0.11. This leads us to suggest that the alignment mechanism in the PEG/hexanol media is, in fact, different from the simple steric model postulated in PALES.

2.5 Discussion

Although the results about dSH3 in Figure 2.7A are quite unexpected, which indicates that nearly-spherical dSH3 domain aligns stronger than the markedly anisotropic dSH3-*sl*-dSH3 construct, however, such results appear paradoxical only in the light of the simple steric model involving planar bicelles. Alternatively, if the morphology of the media is complex, then a steric mechanism can, in principle, lead to an arbitrary alignment tensor. Indeed, assume for the sake of argument that anisotropic media contains cavities that can transiently accommodate certain proteins. Assume further that the cavities are shaped such that they form "lock and key" pair with a given protein. Clearly, the resulting alignment tensor will be different from the one that is predicted by PALES.

Aside from our study, there are also many other literature reports of RDC measurements in the PEG/hexanol media. For model protein ubiquitin, the orientation of the alignment frame is predicted by PALES with reasonable accuracy, NSP = 0.78–0.86. However, the GDO values are overestimated by a factor 2.1–4.5 [51,267]. Our own measurements on ubiquitin in PEG/hexanol media at pH 3.5 led to similar results, with GDO overestimated by 2.1–2.3. The agreement becomes even poorer if the results are scaled according to $\Delta(^2\mathrm{H})$. The data from another small protein, GB3 domain, produce NSP = 0.94 and another poor GDO prediction, off by a factor 3.6 [259]. Taken together with our current data for α-spectrin SH3, these results suggest that PALES can both under- and overestimate the degree of alignment by a wide margin.

This shortcoming seems to be specific to the PEG/hexanol media. In the case of DMPC/DHPC bicelles, the steric PALES model fares much better. For ubiquitin, NSP = 0.96–0.97 and the predicted GDO value is only 1.3 times higher than the experimental value. For GB3 domain, NSP = 0.90 and the predicted GDO is essentially exact. Clearly, it would have been desirable to repeat our experiments in the DMPC/DHPC bicelles. Unfortunately, fSH3 and especially dSH3 constructs tend to precipitate at pH 6.5–7.0 as required by this alignment media[268].

In fact, quite a few studies have shown that the topology of the most used liquid-crystal media is far more complex than previously believed. The DMPC/DHPC bicelles, which were originally described as large disks, have been instead visualized as "two-dimensional networks of attened, highly branched, cylindrical micelles and lamellar sheets perforated by large irregular holes" [269]. Similar observations have been made with regard to the PEG/hexanol media. Early on it has been noted that the L_α bilayers display collective wave-like modulation[270,271]. Besides, it has also been pointed out that dilute conditions used in NMR experiments bring the system close to the transitions region where the lamellar phase L_α coexists with a "sponge phase" L_3[266], and the diffusion data further suggest that the motion of probe molecules between the oriented bilayers (i.e., along the z-axis in the NMR tube) is significantly obstructed. These obstructions may arise from the "bridges" between the adjacent bilayers, the presence of small globular structures (e.g., micelles) in the interbilayer space, etc. None of this potentially complex topology is taken into consideration in the PALES model—which may explain its apparent lack of accuracy, as indicated by Figure 2.7.

From a more general perspective, the failure of PALES can be attributed to site-specific interactions between the proteins and the media. Generally, PEG-based media are protein-friendly in a sense that there is little protein adsorption. Nevertheless, specific interactions involving PEG/hexanol have been observed for short peptides[272], N-terminal domain of the enzyme I from bacterial phosphotransferase system[273], two cytoplasmic domains of chitobiose transporter[274], and folding intermediate of FF do-

main from HYPA/FBP11[275]. Besides, there could also be many cases remaining undocumented but dismissed due to the unsuccessful attempts to prepare aligned samples.

PEG doped with hexanol forms a bilayer where the surface is lined with hydroxyl headgroups. The interactions of the proteins with such bilayers have been modeled in several influential theoretical studies[276,277]. All-atom force field simulations have also been reported[278,279]. These simulations clearly identify the preferred orientation of the protein (lysozyme) on a surface of a PEG bilayer. Of note, the binding interface appears to be formed by charged/polar residues interacting with PEG oxygens at or near the surface of the bilayer[278]. Of course, the binding is only weakly specific since many sites on the protein surface can form such favorable contacts. Note that electrostatic effects play a prominent role in molecular alignment even when the media is not charged—it is sufficient that the media is polar. These aspects has been extensively discussed in several studies based on small molecules[280,281].

To elucidate a possible role of electrostatics in PEG/hexanol media, an additional sample of fSH3 containing 100 mM NaCl has been prepared. If the alignment is purely steric then one would not expect to observe any significant changes upon addition of NaCl. In reality, however, the GDO parameter jumps from 1.19×10^{-3} to 1.85×10^{-3} (see Table 2.1). This result is quite unexpected and, at a first glance, counterintuitive. Indeed, if electrostatic interactions are involved, then addition of salt should screen out these interactions and thus lower the degree of alignment. However, a number of plausible explanations can be suggested for this finding, and the increase in GDO may result from the interplay between the electrostatic and steric mechanisms. Furthermore, it may also occur in the context of purely electrostatic alignment. One possible explanation for such effect is that the excessive electrostatic repulsion drives the solute away from the charged media and into the bulk, where it cannot effectively align[253]. Adding NaCl to the solution reverses this process, thus increasing the degree of alignment. Details of alignment notwithstanding, the

observed dependence of the GDO on the ionic strength of the solution suggests that an electrostatic mechanism plays a role.

To further probe the relevance of electrostatic interactions, an additional series of PALES simulations have been carried out. Unlike the steric model which has no adjustable parameters, the electrostatic model contains several such quantities. The ionic strength of the solution was set to the actual experimental value, 17 mM. The other two parameters, pH and the surface charge density of the medium σ, were optimized in an ad hoc fashion to obtain the best possible agreement with several experimental GDO values. The optimization was conducted by means of the grid-search in the space of two parameters.

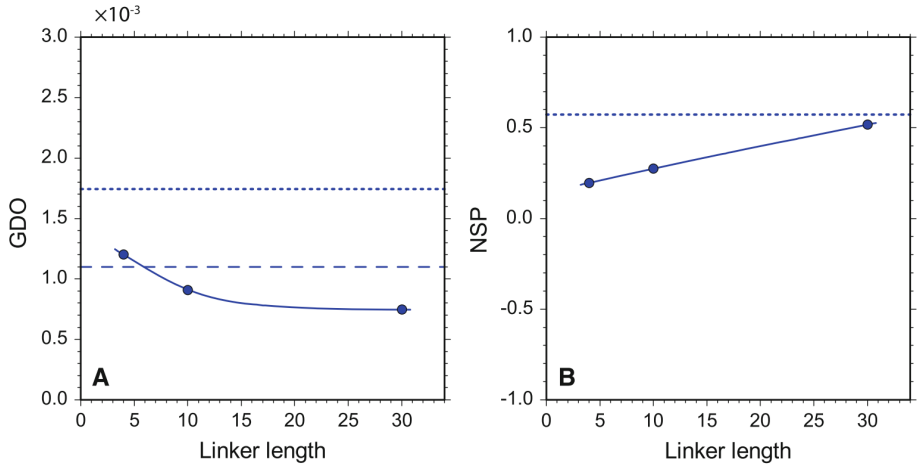


Figure 2.8. The GDO and NSP alignment parameters for α-spc SH3 constructs as obtained from PALES-based electrostatic simulations. The optimized model assumes planar bicelle bilayers with surface charge density $\sigma = 0.05$ e/nm^2, and effective sample pH 4.8.

The optimized electrostatic model produced the predictions for GDO and NSP as shown in Figure 2.8. The main result of this simulation is that it is indeed possible to recreate the situation where dSH3 and fSH3 display a higher degree of alignment than tandem constructs, as observed experimentally (see Figure 2.7A). The parameters

obtained from the model optimization are $\sigma = 0.05$ e/nm^2, effective pH 4.8. The value 0.05 is relatively low in comparison, for example, with Pf1 phage, -0.47 e/nm^2. In principle, surface charge may arise from adsorption of ions on the interface between water and non-ionic surfactants[281–283]. However, it appears that the charge density on the surface of a C12E5 bilayer in water is actually much lower than 0.05 e/nm^2 and negative[283,284], even in the presence of a small amount of ionic surfactant[285,286]. At the same time, the obtained effective pH value, 4.8 is significantly higher than the actual experimental pH, 3.5.

In judging the results of these simulations one should keep in mind that the model of electrostatic alignment is rather crude. Specifically, the pK_a values used in the calculations have a typical error of 0.5–1 unit, with maximum error of up to 2 units[250]. Furthermore, the predictions of electrostatic alignment are highly sensitive to fine details of the charge distribution. For instance, the predicted alignment parameters may change in response to a conformational jump of a single charged side chain. To demonstrate this effect, SH3 conformational ensembles using the BAAAA structure of the SH3 domain have also been generated, which differ from AAAAA structure in that the single side chain, E17, is placed in an alternative conformation. This minor structural modification led to significant decrease in the predicted GDO of the fSH3 sample, from 1.74×10^{-3} to 1.36×10^{-3}.

In conclusion, the presented PALES-based electrostatic simulation demonstrates the feasibility of the scenario in which dSH3 and fSH3 align more efficiently than the tandem constructs. Beyond that, however, it is difficult to find any particular significance according to the results as shown in Figure 2.8. It can be envisaged that more sophisticated prediction tools will be developed in future, where a fully atomistic approach will be extended to the liquid crystal media, for example, there have already been several MD-based studies carried out for studying such systems[278,280,281].

2.6 Conclusion

In this study we focus on proteins comprised of two modular domains connected through a random-coil linker (α-spc SH3 domains connected through glycine-serine linker). When the linker is short, the tandem SH3 construct represents a dumbbell-shaped molecule with relatively little domain-domain mobility. Being placed in steric alignment media, such as PEG/hexanol, this molecule is expected to display a high degree of alignment. As the length of the linker is increased, the domains become effectively uncoupled and start behaving as independent entities. Consequently, the degree of alignment observed in each of the tandem domains approaches that of the isolated SH3 domain.

In qualitative terms, it has been noted before that differential domain alignment conveys information about interdomain mobility[287–290]. Here we attempted to quantify the dependence of the alignment parameters (i.e., amplitude and orientation) on the length of the linker. The form of the corresponding transition curves depends on the details of the system, e.g., the shape of the individual domains and the linker attachment points. In the case of electrostatic alignment, the trend can be easily reversed, i.e., the degree of alignment may increase with lengthening of the linker. The transition phenomenon itself, however, is sufficiently general.

To translate the proposed intuitive model into quantitative predictions, we generated a series of conformational ensembles representing α-spc SH3 constructs with different linker lengths (the disordered linkers and terminal sequences were built using the program TraDES). These ensembles were used to predict the parameters of steric alignment with program PALES. To test the validity of the theoretical predictions, we prepared three tandem constructs of α-spc SH3 and two variants of the isolated SH3 domain. These chimeras were used to measure backbone $^{1}H^{N}$–^{15}N RDCs in two series of PEG/hexanol samples ($r = 0.85$ and 0.96). It was found that the alignment of tandem proteins is in line with expectations and agrees reasonably well with the results from PALES-based simulations. The isolated domains, however, displayed much stronger alignment that is much expected. Furthermore, the degree of align-

ment turned out to be sensitive to the ionic strength of the solution, which is generally not expected to be the case for a well-behaved steric alignment media.

Several factors may potentially contribute to this situation. First, the topology of the liquid crystal media is known to be more complex than envisaged in the simple PALES model (i.e., planar bilayers). Furthermore, there is a possibility that the phase equilibrium of the liquid crystal changes in response to addition of protein or salt (these changes may not necessarily be fully reflected in pK_a)[248]. It is also likely that protein alignment in PEG/hexanol significantly depends on weak site-specific interactions. In particular, an electrostatic mechanism appears to play a role, visibly affecting the alignment of isolated domains.

In summary, it seems that PEG/hexanol bears little resemblance to an idealized alignment media envisioned in PALES, therefore special attention should be paid on it especially in the context of studies where PALES is used to simulate RDC data from disordered proteins dissolved in PEG-based media[291–293]. Indeed, if the results cannot be reliably predicted about the alignment for globular proteins, the predictions should be even less reliable for disordered proteins. In this latter case, all residues can potentially form site-specific contacts with the media and, furthermore, these contacts may significantly skew the original conformational equilibrium (for those protein molecules that are in the vicinity of the media and give rise to observable RDCs).

The chimera protein investigated in this work is comprised of two non-interacting domains connected by a variable-length random linker. It can thus be described as an example of the system with minimal domain coupling. This example provides an important point of reference for future studies of the proteins with substantial domain coupling, either through domain–domain interactions (e.g., Crk adaptor protein[210], Syk kinase[9]) or through a structured linker (e.g., troponin[294]). In the case of popular PEG/hexanol media, the usefulness of this approach is presently limited by lack of precise information about the morphology of the oriented phase and the details of alignment mechanism. It is expected that the DMPC/DHPC media may offer a better

chance of success, although it is far more restrictive in terms of sample conditions. Further progress in this area depends on development of more accurate alignment prediction tools. Ultimately, such tools should be based on all-atom MD simulations to accurately model the interactions between the protein and the alignment media.

2.7 Major Contribution into This Thesis by Other Investigators

Ryan Muir made some contributions to the early version of the program for two-domain protein *in silico* structural model generation. Etti Harms provided helpful suggestions about primer design strategies for tandem α-spc SH3 proteins. Nina Gorenstein, Josh Ward and Yi Xue provided some useful ideas or comments about the design of experiments.

CHAPTER 3. NMR PULSE SEQUENCES DEVELOPMENT FOR DISORDERED PROTEIN STUDIES

3.1 ^{15}N R_2 Measurement for Disordered Proteins with PD-CPMG

3.1.1 Introduction

^{15}N relaxation measurement is one of the most useful NMR experiments—it is fast, reasonably simple to interpret, and immediately provides information on protein backbone flexibility. This information can be further used to characterize the role of protein motions (i.e., conformational entropy) in ligand binding[295,296], protein stability[297,298], allosteric regulation[299,300] etc.

The spin context of ^{15}N relaxation is fairly straightforward. It is normally assumed that nitrogen relaxation is driven by $^{1}H^{N}$–^{15}N dipole–dipole (DD) interaction and, to a lesser degree, by ^{15}N chemical shift anisotropy (CSA) mechanism. Nevertheless, even in this sufficiently simple spin system there is room for uncertainty—the appropriate length of N–H bond was a subject of debate and the degree of anisotropy and site-to-site variation of the ^{15}N CSA tensors have also been widely discussed[301,302]. Furthermore, small and difficult-to-quantify R_{ex} contributions into ^{15}N R_2 appear to be more common than previously thought[303].

In addition to these fundamental sources of uncertainty, there are also a number of errors associated with specific measurement schemes that have been uncovered in recent years. For example, it has been found that frequent application of ^{1}H pulses during the spin-lock period with the intention to suppress dipolar–CSA cross-correlation may actually re-introduce the undesirable J-coupling evolution[135]. In ^{1}H–^{15}N saturation-transfer NOE experiment it was found that the standard saturation

sequence tends to create a number of unwanted spin modes[304]. Besides, there are quite a few artifacts caused by the solvent exchange effects for relaxation measurement[305].

Finally, there are also uncertainties associated with the choice of the model. The most common approach is to choose from several versions of the model-free approaches which make different assumptions about the underlying dynamics—in particular, with regard to local motions on $\sim$ 1 ns time scale[306]. While there are statistical criteria that guide this choice, it has been shown that the outcome is often influenced by the random noise-like errors and depends on availability of data measured at different magnetic fields[307]. An additional level of complexity is added by the proteins overall tumbling, which should be properly modeled using a fully asymmetric diffusion tensor[308,309].

With all of these sources of uncertainty combined, ^{15}N-relaxation-derived motional parameters are not necessarily very accurate. Consider, for example, the well-studied model protein ubiquitin, where several groups independently measured and analyzed ^{15}N relaxation in solution[193,310,311]. The backbone order parameters S^2 reported in these studies differ significantly from each other: the rms deviations between the independently determined sets of S^2 are in the range from 0.02 to 0.07 and the correlation coefficients are in the range from 0.56 to 0.94.

In practice, however, the perceived lack of accuracy is not very important so long as most experimental studies focus on the difference in order parameters, ΔS^2, rather than the absolute values. It is reasonably safe to assume that systematic biases subtract out in calculating ΔS^2 and the result correctly represents changes in local dynamics. The examples include changes in order parameters in response to ligand binding, point mutations, various post-translational modifications, etc.[312,313]. Given that such changes are typically small, on the order of 0.05 or less, the primary concern is the precision of the data—it is important to ensure that small ΔS^2 values can be experimentally reproduced and quantitated in a reliable fashion.

These considerations are especially relevant for intrinsically disordered proteins (IDPs), which have emerged as the area of major biomedical interest[69]. IDPs do not

lend themselves to the standard methods of structural characterization. Instead, the dynamics-oriented experiments play a prominent role. In particular, ^{15}N relaxation and paramagnetic relaxation enhancements (PREs) are arguably two most informative experiments in the field of intrinsically disordered proteins. At the same time, it is often considerably more difficult to achieve high precision in measuring ^{15}N relaxation rates for IDPs. Many disordered proteins are prone to aggregation which makes it necessary to work with low-concentration NMR samples, often in the range 10–100 μM, besides, data collection has to be carried over a short period of time in order to avoid sample degradation. If the studies are conducted at or near physiological conditions (37 °C, pH 7.4), the quality of the already crowded spectra suffers from the line-broadening caused by solvent exchange k_{ex}, which further lowers the signal-to-noise ratio in NMR spectra.

For disordered proteins at 37 °C, the ^{15}N T_2 relaxation times are relatively long, on the order of 0.5–1.0 s. On the other hand, the length of the spin-lock period τ_{rel} in the standard $R_{1\rho}$ experiment does not exceed $\sim$ 100–150 ms, as dictated by hardware limitations. It is clearly not possible to properly sample the magnetization decay curve over such a short interval of time. This situation is illustrated in Figure 3.1A, which shows the $R_{1\rho}$ decay curve from residue L15 in a 100 μM sample of denatured ubiquitin at 37 °C. The expansion of this relaxation profile in Figure 3.1B demonstrates the problem with such measurement—generally, when τ_{rel} is much shorter than T_2 the decay profile appears flat, which makes it difficult to accurately determine the relaxation rate constant. A rigorous treatment indicates that the best results are obtained when τ_{rel} is extended beyond T_2[137].

The restrictions concerning the maximum length of τ_{rel} are ultimately due to low gyromagnetic ratio of nitrogen, according to $\omega_1 = \gamma_N B_1$. In order to generate sufficiently strong ^{15}N *rf* field which is necessary to suppress the effect of $^1J_{NH}$ coupling, a very strong current should be applied to the transmitter coil. This pushes the probe to the limits of its performance. In the case of spin-lock measurements, the room-temperature probe can sustain 1.7 kHz nitrogen *rf* field for a period of time $\sim$

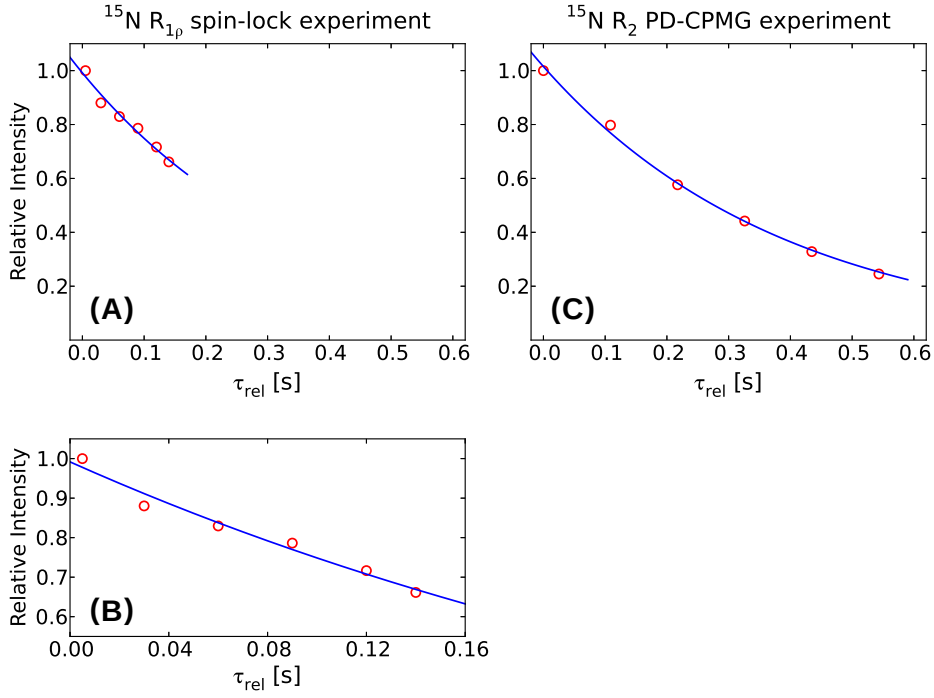


Figure 3.1. Examples of relaxation curves from ^{15}N $R_{1\rho}$ spin-lock experiment (panels A and B) and R_2 PD-CPMG experiment (panel C). The data are from residue L15 in 100 μM sample of unfolded ubiquitin at 37 °C. Panel (B) is the expansion of panel (A).

150 ms. In the case of R_2 CPMG measurements, the probe can withstand the pulse train with the repetition time of 1 ms for $\sim$ 100 ms. For early-generation cryogenic probes the power limits are even more restrictive. Extending τ_{rel} beyond these limits can lead to excessive heating of the sample and cause probe arcing which severely degrades the performance of the pulse sequence[314], or the probe can be permanently damaged in the worst scenario.

Such issue can be overcome by using more advanced hardware, and it appears that the latest-generation hardware and particularly new cryogenic probes should make it possible to run ^{15}N spin-lock experiment with much longer τ_{rel}[315], however, the availability of such advanced hardware is still quite limited. There are also several strategies developed that allow overcoming this problem without recourse to advanced hardware—it has been demonstrated that $^{1}J_{NH}$ coupling can be suppressed by using

relatively weak ^{15}N spin lock plus a small number of strategically placed ^{1}H pulses[316], however, this is semi-selective and in order to collect a complete $R_{1\rho}$ dataset it is necessary to repeat the measurements several times using different ^{15}N carrier settings.

The alternative solution that has been used in this work, is to apply ^{1}H decoupling to suppress $^{1}J_{NH}$. Because of the high gyromagnetic ratio of the proton ($\gamma_H \simeq 10\gamma_N$), it is easy to generate a sufficiently strong *rf* field without overloading the probe ($\omega_1 = \gamma_H B_1$). In the presence of proton decoupling, ^{15}N 180° pulses can be applied infrequently only for the purpose of refocusing chemical shift evolution. Using this approach, τ_{rel} can be extended by several times compared with most traditional schemes without risking adverse consequences to a probe.

The pulse sequence based on such idea, proton-decoupled CPMG (PD-CPMG), has been introduced in several previous work for ^{15}N relaxation measurement for proteins[317–319]. However, in all of these reports it is noted that there exists certain amount of bias for ^{15}N R_2 rates measured with such scheme, on the order of 5%, and because of this unresolved issue PD-CPMG has found little use during the last two decades.

In this work we have developed a new and upgraded version of PD-CPMG which can reliably reproduce the results of the well-established $R_{1\rho}$ spin-lock experiment. The new pulse sequence allows one to extend the observation window τ_{rel} to 500 ms and beyond (Figure 3.1C). The proper sampling of the relaxation decay curve, down to the level of 0.2 on the relative intensity scale, facilitates the accurate determination of R_2. For the samples at hand, the precision of the relaxation measurements is improved by up to a factor of 3 compared to the standard $R_{1\rho}$ experiment. In addition to the extended observation window, the new version of PD-CPMG experiment also offers many other valuable properties such as more easily to perform ^{15}N offset corrections and more robust to artifacts caused by solvent exchange effects.

3.1.2 Materials and Methods

SH3 domain from chicken α-spectrin and human ubiquitin have been expressed in ^{15}N-enriched M9 media and purified with previously developed protocols[256,320]. Protein concentrations were determined using UV absorbance at 280 nm. The sample of 1.5 mM α-spc SH3 has been prepared in H_2O, 20 mM citrate, 0.02% NaN_3, pH 3.5. The sample of 2.0 mM unfolded ubiquitin was prepared in H_2O, 8 M urea, pH 3.0. Two additional ubiquitin samples were prepared with protein concentration 50 and 100 μM. Sample volume in each case was 500 μL, with 100 μL of D_2O added into the coaxial insert of the NMR tube (Wilmad, WGS-5BL) to serve as external lock.

All NMR experiments were performed on Varian Inova 600 MHz spectrometer equipped with triple-resonance z-axis gradient probe. The measurements were carried out at 25 °C for α-spc SH3 and 37 °C for unfolded ubiquitin. The backbone amide spectra were recorded with ^{15}N carrier at 119 ppm and ^{15}N spectral width of 30 ppm (α-spc SH3) or 22 ppm (ubiquitin). The data were collected with 64 t_1 increments and 1800 points in t_2 domain. The spectra of Arg ($^1H^{\varepsilon}$,$^{15}N^{\varepsilon}$) in α-spc SH3 were recorded with ^{15}N carrier as 85 ppm, ^{15}N spectral width of 10 ppm, and 16 t_1 increments. All data were collected with 4 scans per t_1 increment. The spectra were processed using NMRPipe[261]. Squared phase-shifted sine-bell window function was applied in both dimensions and the spectra were zero-filled to the size 4096×512 ($^1H^N$,^{15}N) or 4096×256 ($^1H^{\varepsilon}$,$^{15}N^{\varepsilon}$).

To ensure a clean comparison between PD-CPMG and spin-lock[135] experiments, we have amended the standard $R_{1\rho}$ pulse sequence by adding a water-crusher element[321]. In the spin-lock experiment we have used the following τ_{rel} delays: 5(1), 30(1), 60(2), 90(3), 120(4), and 140(4) ms (shown in brackets is the number of ^{1}H 180° pulses used to suppress DD/CSA cross-correlations). The ^{15}N spin-lock has been applied with *rf* field strength 1.7 kHz. All other experimental settings were the same as in PD-CPMG experiment, including the recycling delay $d_1 = 2$ s. ^{15}N R_1 data have also been recorded in order to determine ^{15}N offset correction for both $R_{1\rho}$ and PD-CPMG data.

The spectra were integrated using the autoFit/nlinLS subroutines from the NMRPipe package. These subroutines operate on the entire "stack" of spectral planes, making an assumption that the peak position and the linewidths do not change significantly from one plane to the other and hence can be fitted in a global sense. This approach is especially helpful for quantitation of the weak peaks. The data were fitted assuming Gaussian peak shapes, as appropriate for the squared sine-bell apodization. The spectral assignments of α-spc SH3[238] and denatured ubiquitin[322] were obtained from the previous reports.

To monitor sample temperature during the PD-CPMG experiment we have developed a special scheme based on the pulse sequence as shown in Figure 3.2. Starting from the data acquisition point the pulse sequence is modified as: g_{11} – τ_{blank} – $p_{H(90)}$ – t_2 – d_1. Here g_{11} is the clean-up gradient with the strength 20 G/cm and duration 3.5 ms, τ_{blank} is 100 ms delay during which time the ^{15}N WALTZ-16 decoupling is turned on, $p_{H(90)}$ is hard proton pulse with the same phase as the receiver, t_2 is the acquisition period of 0.5 s, and d_1 is the recycling delay of 1.5 s. The purpose of this sequence is to record 1D spectrum of methanol (which is contained in the NMR tube coaxial insert) under the conditions that emulate the actual PD-CPMG measurement.

We have simulated spin evolution of the two-spin ($^1H^N$,^{15}N) and three-spin ($^1H^N$, ^{15}N and $^1H^\alpha$) systems during the τ_{rel} period of the PD-CPMG and spin-lock experiments. Taken into consideration were all sources of coherent evolution: *rf* fields, frequency offsets, and scalar couplings. In addition, Redfield matrix has been generated by rigorously including DD and CSA contributions. In the 2-spin simulations we have also included the "external" $^1H^\alpha$ spin, which manifested itself through dipolar contribution into the auto-relaxation rates of ($^1H^N$,^{15}N) spin modes. In the 3-spin simulations, $^1H^\alpha$ was treated as a part of the spin system. We have assumed that $^3J_{HNH\alpha} = 10$ Hz, $^2J_{NH\alpha} = 5$ Hz which is the upper limit for the respective experimentally observed couplings[323,324]. ^{15}N and $^1H^N$ CSA tensors were modeled according to the literature data[302,325]; $^1H^\alpha$ CSA tensor was assumed to be zero. The distance between $^1H^N$ and $^1H^\alpha$ was set to 1.84 Å. In order to emphasize the effect of DD/DD

cross-correlations, the angle between ^{15}N–^{1}H^{N} and ^{1}H^{N}–^{1}H$^{\alpha}$ vectors was set to zero. All simulations used a simple form of Lipari–Szabo spectral density[18], where it was assumed that $\tau_R = 2$ ns, $S^2 = 0.5$, and $\tau_{\text{fast}} = 0$. This parameterization is consistent with the previous analyses of ^{15}N relaxation in denatured protein[326] and correctly reproduces typical R_1 and R_2 values observed in denatured ubiquitin at 37 °C. *Rf* pulses were simulated with finite length. To model the effect of *rf* field inhomogeneity, the simulations were repeated for 100 discrete values of *rf* field strength uniformly distributed in the interval $[\nu_1 - 2\sigma, \nu_1 + 2\sigma]$. The signal intensity was then calculated as a weighted average of the results from multiple simulations. Solvent exchange was modeled as described previously[327]. Specifically, any multi-spin mode combining ^{15}N and ^{1}H^{N} spin operators undergoes exchange-induced decay with the rate as k_{ex}.

The simulations featuring 15×15 and 63×63 propagator matrices were repeated for the same set of τ_{rel} delays as used experimentally. The simulated signal intensities were then fitted with a single exponential, yielding the values of $R_{1\rho}$ or R_2^{app}. However, it turns out that results from 2- and 3-spin simulations are virtually identical, which can be well understood considering that: (i) DIPSI-2 sequence has been designed to perform well in the presence of homonuclear couplings, such as $^{3}J_{\text{HNH}\alpha}$, (ii) the train of ^{15}N 180° CPMG pulses is sufficient to eliminate the effect of the small $^{2}J_{\text{NH}\alpha}$ coupling, with off-resonance DIPSI-2 providing additional insurance, and (iii) three-spin modes generated by DD/DD cross-correlations are effectively dephased by DIPSI-2 decoupling. More details about the numerical simulation setup for PD-CPMG are included in Appendix B.

3.1.3 Results

3.1.3.1 PD-CPMG Pulse Sequence for ^{15}N R_2 Measurements

The new pulse sequence shown in Figure 3.2 is essentially a decoupled, sensitivity-enhanced HSQC, similar to what has been previously used for ^{15}N $R_{1\rho}$ measurements[135]. Inserted in this sequence is the proton-decoupled CPMG element of the

net τ_{rel}. The CPMG pulse train uses [0013] phase cycle which was shown to have favorable compensation properties[328–331]. The pulse repetition rate in the PD-CPMG sequence is relatively low, with the pulse spacing $2\tau_{cp} \sim 14$ ms. The proton decoupling during the CPMG train is achieved by application of the DIPSI-2 sequence[138] with the *rf* field strength 4 kHz. The relatively low level of combined ^{1}H and ^{15}N *rf* power allows τ_{rel} period extended to 500 ms and beyond.

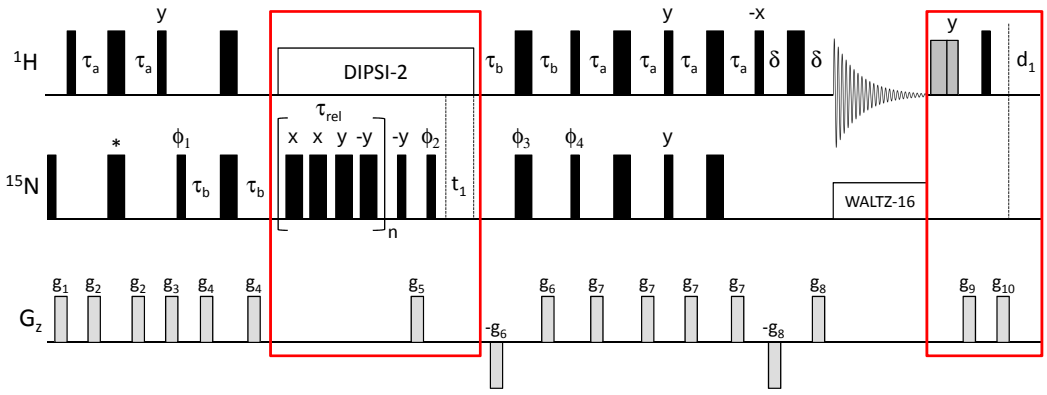


Figure 3.2. PD-CPMG pulse sequence for measuring ^{15}N R_2 relaxation rates. The design is mainly based on the $R_{1\rho}$ spin-lock pulse sequence[135], and the portions of the sequence which are new and distinct are enclosed in red boxes. Two consecutive water purge pulses have the durations of 6.0 and 3.7 ms and use the *rf* field of 12.4 kHz (shaded rectangles in the plot). Gradient strengths in G/cm (length in ms) are $g_1 = 5.0$ (1.0), $g_2 = 4.0$ (0.5), $g_3 = 10.0$ (1.0), $g_4 = 8.0$ (0.5), $g_5 = 25.0$ (1.0), $g_6 = 15.0$ (1.25), $g_7 = 4.0$ (0.5), $g_8 = 14.8$ (0.125), $g_9 = 20.0$ (3.5), $g_{10} = 20.0$ (2.0). The phase cycle employed is $\phi_1 = (x, -x)$, $\phi_2 = y$, $\phi_3 = 2(x), 2(y), 2(-x), 2(-y)$, $\phi_4 = x$, $\phi_{rec} = (x, -x, -x, x)$. DIPSI-2 has initial *rf* field applied along x. For arginine side-chain $^{15}N^{\varepsilon}$ R_2 measurement, the rectangular ^{15}N 180° pulse marked by (*) should be replaced with r-SNOB pulse[332] with ~ 15 ppm bandwidth, and ^{15}N 180° pulses in the CPMG train need to be changed into REBURP shapes[333] with ~ 15 ppm bandwidth covering $^{15}N^{\varepsilon}$ resonances, but not $^{15}N^{\eta}$.

In the pulse sequence Figure 3.2, proton decoupling is extended to also cover t_1 evolution period. It is well known that proton decoupling during t_1 generally leads to

sharper spectral peaks[334,335]. This is particularly important for IDPs where the spectra are highly crowded. Under physiological conditions, IDPs experience fast solvent exchange which efficiently destroys two-spin modes such as $2N_xH_z$[327]. For conventional HSQC experiment, this causes substantial line broadening in ^{15}N dimension resulting in poor-quality spectra. In contrast, decoupled HSQC experiment records the evolution of pure in-phase coherence, N_x, and therefore avoids this problem.

In order to obtain accurate measurement results for ^{15}N R_2 rates with PD-CPMG, the ^{15}N 180° CPMG pulses should be calibrated with care, which can be performed with HSQC-based calibration scheme. The calibration of ^{1}H 90° pulse for DIPSI-2 decoupling does not need to be particularly accurate since PD-CPMG sequence is highly tolerant to pulse calibration errors on ^{1}H. However, it is important that the duration of τ_{cp} should be calculated rigorously according to the synchronization condition and always corresponds to integer number of DIPSI-2 cycles, see Equation (3.2).

Note that fast solvent exchange partially cancels the advantage of the sensitivity enhancement in the pulse sequence Figure 3.2. A maximum potential gain from the use of the sensitivity enhancement scheme is given by a factor of $\sqrt{2}$[336]. However, the sequence requires an extra transfer step of $\sim$ 5 ms duration. Assuming that amide solvent exchange rate is 100 s^{-1} , it is easy to estimate that the loss of signal due to solvent exchange completely offsets the sensitivity gain. Nevertheless, for most samples of practical interest the exchange rate is lower than 100 s^{-1} and therefore the use of the sensitivity enhancement scheme is likely justified, besides, enhanced sensitivity element also ensures better water suppression properties.

Finally, under the conditions of fast solvent exchange there is a potential risk associated with partial saturation of the water signal[337,338]. Generally speaking, the degree of saturation is variable and, in particular, depends on τ_{rel}. Since the saturation is effectively transmitted to the amide sites, the magnitude of $^{1}H^{N}$ magnetization available prior to each scan also turns out to be dependent on τ_{rel}. This causes an unwanted modulation of the measured relaxation curves, which can seriously compro-

mise the accuracy of the experiment[339,340]. To avoid these artifacts we have chosen to "crush" water magnetization following the acquisition period and prior to the recycling delay, thus creating highly reproducible initial conditions. This design leads to a very accurate experiment. In principle, the sensitivity is to a certain degree sacrificed relative to a sequence that preserves water magnetization. However, in practice the sensitivity is similar to the standard ^{15}N relaxation experiments where the water signal is strongly saturated with no regard for solvent exchange.

3.1.3.2 Results from PD-CPMG are Consistent with ^{15}N Spin-lock Experiment

In order to confirm that PD-CPMG scheme achieves consistent results with other commonly used R_2 measurement strategies such as ^{15}N spin-lock experiment, we have collected both types of data on the samples of α-spectrin SH3 domain (1.5 mM) and unfolded ubiquitin (2.0 mM). Only those peaks that do not have any significant spectral overlaps have been selected for analyses (58 residues for α-spc SH3 domain and 48 residues for unfolded ubiquitin). The quality of the relaxation curves is excellent in both experiments, with very low fitting residual. Importantly, the results of the two independent measurements show an excellent correlation with each other, as evidenced by the correlation coefficients 0.999 and 0.993, see Figure 3.3. The systematic deviation between the two datasets is very small. Specifically, the difference between the mean values of R_2 from the new PD-CPMG experiment and the standard spin-lock experiment is $+0.9\%$ in the case of α-spc SH3 and -1.3% in the case of ubiquitin.

We have also recorded a control PD-CPMG experiment where the strength of the DIPSI-2 decoupling was lowered from $\sim$ 4 kHz to $\sim$ 3 kHz. The results remain virtually unchanged, as illustrated in Figure 3.4 (correlation coefficients 0.999 and 0.998). The observed amount of systematic deviation is vanishingly small, therefore it means that PD-CPMG experiment with relatively low level of ^{1}H decoupling power, $\sim$ 3 kHz, can be used to sample the relaxation profiles up to $\tau_{\rm rel} \sim 1$ s and beyond.

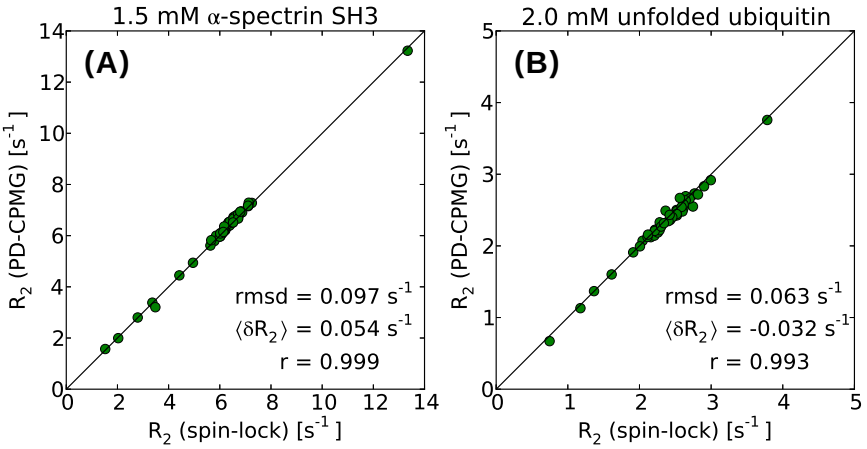


Figure 3.3. Comparison of ^{15}N R_2 rates from the PD-CPMG experiment and the standard spin-lock measurement. The PD-CPMG experiment used $\sim$ 4 kHz DIPSI-2 decoupling. All datasets have been corrected for ^{15}N frequency offset effects. PD-CPMG experiment on the sample of α-spc SH3 was carried out with $2\tau_{cp}$ = 13.7373 ms and the set of six τ_{rel} delays ranging from 0 to 275 ms.

It can also be used to collect the data from samples with high ionic strength where sample heating due to deposition of the *rf* power needs to be reduced.

3.1.3.3 PD-CPMG Scheme Improves the Precision of ^{15}N R_2 Measurements

The PD-CPMG sequence is intended to improve the sampling of ^{15}N relaxation curves in the situation when R_2 relaxation is relatively slow, on the order of 1–3 s^{-1}. Specifically, the experiment is relevant for IDPs at or near the physiological temperature. Of necessity, NMR studies of IDPs are often conducted at low concentration, 100 μM or less. Furthermore, the time frame of the measurements is usually limited, since the samples have a tendency to rapidly deteriorate. For those samples where pH is close to physiological, the signal-to-noise ratio is further lowered by amide solvent exchange which in the case of IDPs affects all residues in the protein. Under these circumstances, the precision of ^{15}N R_2 data can become a critical issue.

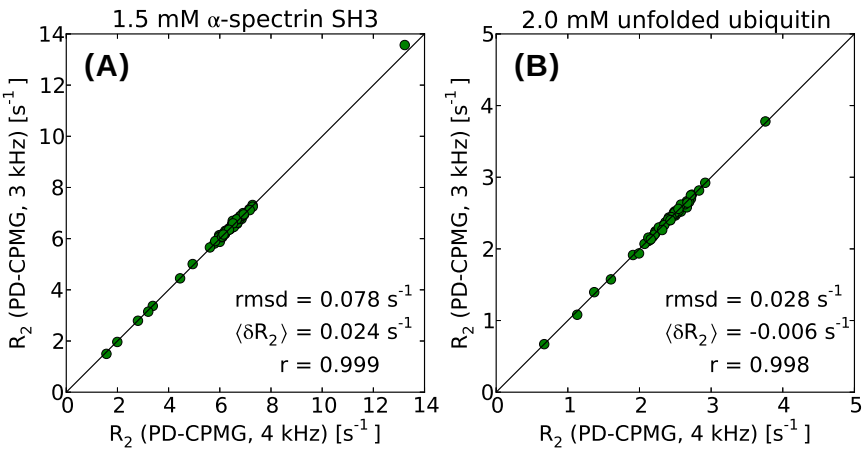


Figure 3.4. Comparison of ^{15}N R_2 rates from the PD-CPMG measurements using $\sim$ 4 kHz and $\sim$ 3 kHz DIPSI-2 decoupling. In the experiments with $\sim$ 3 kHz decoupling, the data from α-spc SH3 were obtained using m = 1, $2\tau_{cp}$ = 9.6509 ms, and the set of six τ_{rel} delays ranging from 0 to 193 ms, whereas the data from ubiquitin were collected using m = 1, $2\tau_{cp}$ = 9.8235 ms and the set of six τ_{rel} delays ranging from 0 to 589 ms.

In order to model this type of a situation, we have prepared two samples of unfolded ubiquitin with protein concentration 100 and 50 μM and collected the relaxation data at 37 °C. Toward this end, we employed the standard ^{15}N $R_{1\rho}$ experiment [135] as well as the new PD-CPMG scheme. All experiments were recorded in duplicate to assess the reproducibility of the results. The outcome is illustrated in Figures 3.5 and 3.6.

The relaxation curves recorded in the repeat spin-lock and PD-CPMG experiments are illustrated in Figure 3.5. The results for residue L15 are shown as is, i.e., in a form of $R_{1\rho}$ decay and apparent R_2^{app} decay, with no correction for ^{15}N frequency offset. The quality of the fits is similar for the two experiments, as can be appreciated from the magnitude of the fitting residual χ. However, the reproducibility of the relaxation rates as determined in the spin-lock measurements (2.41 and 2.81 s^{-1}) is clearly worse than for PD-CPMG (2.67 and 2.56 s^{-1}). This happens because the spin-

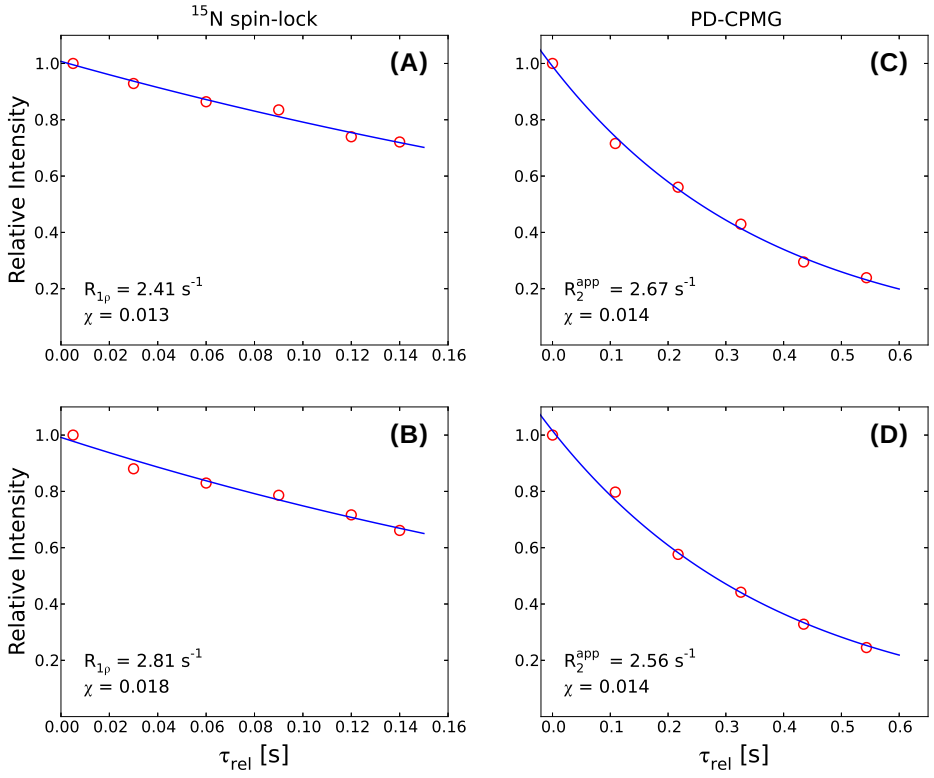


Figure 3.5. Relaxation decay profiles for residue L15 in 0.1 mM unfolded ubiquitin sample as obtained from the repeat measurements using ^{15}N spin-lock experiment (panels A and B) and PD-CPMG (panels C and D). The measurements using 2.0 mM sample of unfolded ubiquitin yield $R_{1\rho} = R_2^{app} = 2.58\ s^{-1}$ for residue L15. Parameter χ is the rmsd (fitting residual) between the experimental and fitted intensities.

lock experiment is limited to short τ_{rel} delays and therefore cannot properly sample the relaxation decay profile[137].

The summary of the data from the repeat experiments is shown in Figure 3.6. Clearly, PD-CPMG achieves a much better reproducibility than the standard spin-lock measurement[135]. The average absolute-value error in $R_{1\rho}$ data is three times as large as in PD-CPMG R_2^{app} data: 9.1% vs. 3.7% for 100 μM sample and 22.6% vs. 7.5% for 50 μM sample. This difference in precision can be critical for studies that use ^{15}N relaxation to quantify changes in IDPs conformational dynamics in

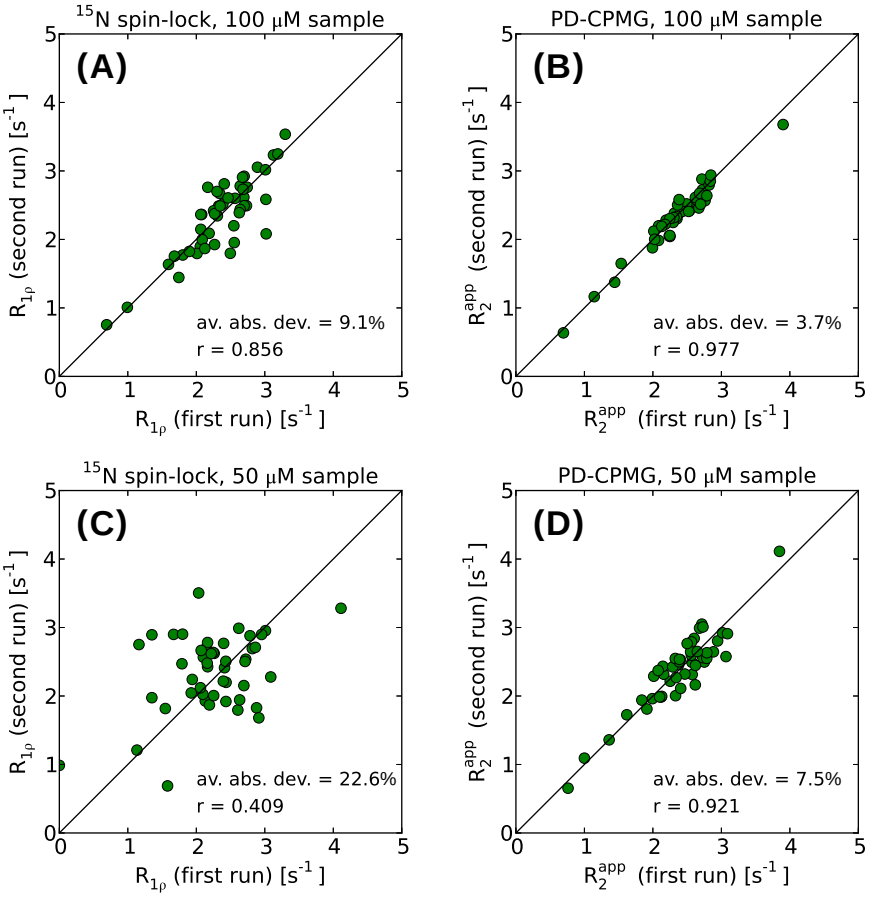


Figure 3.6. The reproducibility of $R_{1\rho}$ and R_2^{app} relaxation rates as measured in the repeat spin-lock and PD-CPMG experiments using low-concentration unfolded ubiquitin samples at 37 °C.

response to various perturbations, e.g., point mutations, posttranslational modifications, variations in pH, crowding effects, etc. Such changes are typically subtle and can be successfully identified only when high-precision ^{15}N relaxation data are available. Note that $R_{1\rho}$ data shown in Figure 3.6C are essentially useless in this regard since the measurement errors are of the same magnitude as per-residue R_2 variations. In contrast, R_2^{app} data shown in Figure 3.6D retain the ability to distinguish between

mobile residues, constrained residues, and those residues that experience exchange broadening.

3.1.3.4 Modified PD-CPMG Scheme for Arginine Side-chain $^{15}N^{\varepsilon}$

After small modifications, the PD-CPMG scheme (Figure 3.2) can also be used to measure R_2 relaxation rates of $^{15}N^{\varepsilon}$ spins in Arg side chains. Arginine side chains are long and flexible: in globular proteins they are typically immersed in solvent and move in a relatively unconstrained fashion. Consequently, $^{15}N^{\varepsilon}$ R_2 rates tend to be low compared to the backbone ^{15}N rates and thus particularly well suited for PD-CPMG measurements. It is also worth noting that N^{ε} are more prone to solvent exchange than the backbone amide sites. Therefore, the favorable properties of the PD-CPMG experiment with regard to solvent exchange are especially useful in the context of $^{15}N^{\varepsilon}$ measurements.

The pulse sequence requires some alterations compared with the PD-CPMG sequence shown in Figure 3.2. The main source of difficulties is the weak $^{15}N^{\varepsilon}$–$^{15}N^{\eta}$ two-bond coupling with the magnitude ~ 1 Hz[341]. This coupling remains active during the standard CPMG pulse train which employs hard nitrogen pulses. Since τ_{rel} times employed in our measurements are long, the $^{2}J_{N\varepsilon N\eta}$ coupling achieves partial conversion of N_x^{ε} into $2N_y^{\varepsilon}N_z^{\eta}$ during this time period, which causes partial loss of signal and distorts the observed relaxation decay profiles.

Since $^{15}N^{\varepsilon}$ and $^{15}N^{\eta}$ have significantly different chemical shifts (separated in the spectrum by ~ 10 ppm), this problem can be easily overcome by application of selective 180° ^{15}N pulses during the CPMG period. For this purpose we have employed REBURP pulses which have been used in the context of CPMG measurements before[342,343]. One of the potential problems with such strategy is that REBURP pulses are relatively long, ~ 4.0 ms, so that nitrogen magnetization spends part of the time away from the transverse plane. As a result, the measured relaxation decay may potentially represent a mixture of R_2 and R_1 rates. However, numerical simulations suggest that this effect introduces only a very small error, on the order of $\sim 1\%$. To

understand this result it is necessary to take several factors into consideration: (i) REBURP pulses are applied relatively infrequently, $2\tau_{cp} \sim 10$ ms. (ii) REBURP pulses are applied on-resonance or close to resonance since $^{15}N^{\varepsilon}$ chemical shifts are confined to a relatively narrow range. Under these conditions the magnetization mostly remains in the transverse plane during the REBURP pulse. (iii) REBURP pulses are applied with [0013] phase cycle. Considering N_x magnetization, it should be locked by the first two pulses within each cycle, such that no R_1 character is introduced into spin relaxation. Similarly N_y magnetization is locked by the last two pulses within each cycle. (iv) In rapidly moving Arg side chains, R_2 and R_1 rates tend to be similar.

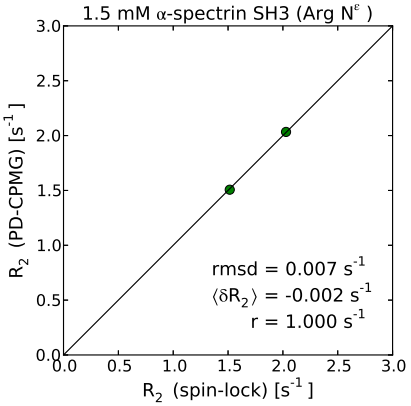


Figure 3.7. Comparison of arginine $^{15}N^{\varepsilon}$ R_2 rates from the PD-CPMG experiment and the standard spin-lock measurement. The results are from two arginine residues in α-spc SH3 domain, R21 and R49.

Ultimately the comparison with $R_{1\rho}$ experiment is necessary in order to confirm that PD-CPMG measurement is error-free. Importantly, $R_{1\rho}$ experiment is insensitive to the effect of $^{2}J_{N\varepsilon N\eta}$. Numerical simulations show that given a relatively large chemical shift offset between $^{15}N^{\varepsilon}$ and $^{15}N^{\eta}$, ~ 10 ppm, moderate spin-lock strength, $\sim$ 1.5–2 kHz, very weak coupling, $\sim$ 1 Hz, and relatively short τ_{rel} delays, $\sim$ 100 ms, it is safe to neglect Hartmann–Hahn transfer during the spin-lock period. The

comparison between the R_2 rates obtained from the standard $R_{1\rho}$ experiment with those measured by $^{15}N^{\varepsilon}$ PD-CPMG is shown in Figure 3.7. The two datasets are consistent within 0.5%, which is approximately the same as the level of agreement previously observed for the backbone amide sites. This result provides validation for the modified version of the PD-CPMG sequence aimed at $^{15}N^{\varepsilon}$ spins. Potential gains in precision in this case should be even more significant than for backbone measurements.

3.1.4 Experimental Details

3.1.4.1 ^{15}N Offset Dependence

One significant shortcoming of the traditional CPMG experiment is that it requires a numerical procedure to correct for ^{15}N offset effect which is not quite straightforward[139]. In contrast, the PD-CPMG experiment using [0013] phase cycle that allows for a simple analytical correction[330]:

$$R_2^{app} \approx R_2 + (R_1 - R_2)\mu^2 \quad (\mu = \nu_{off}^{N}/\nu_1^{N(180)}) \tag{3.1}$$

R_2^{app} is the apparent decay rate as obtained directly from the PD-CPMG measurement, ν_{off}^{N} is the nitrogen frequency offset for a given amide resonance, and $\nu_1^{N(180)}$ is the field strength of the hard ^{15}N 180° CPMG pulses in the units of Hz. For the experimental settings used in this study the magnitude of μ^2 does not exceed 0.03. Consequently, it is safe to ignore higher-order corrections that are proportional to μ^4.

The comparison for ^{15}N offset properties between PD-CPMG and spin-lock experiments is shown in Figure 3.8. In Figure 3.8A it illustrates the dependence of the apparent R_2^{app} rate on ^{15}N frequency offset ν_{off}^{N}. Shown are the simulated R_2^{app} rates (red circles) as well as the predictions using Equation (3.1) (black curve). The true value of R_2 is indicated by the horizontal dashed line. This result suggests that the rates measured via PD-CPMG experiment and corrected by means of Equation (3.1) provide a very accurate measure of R_2 relaxation.

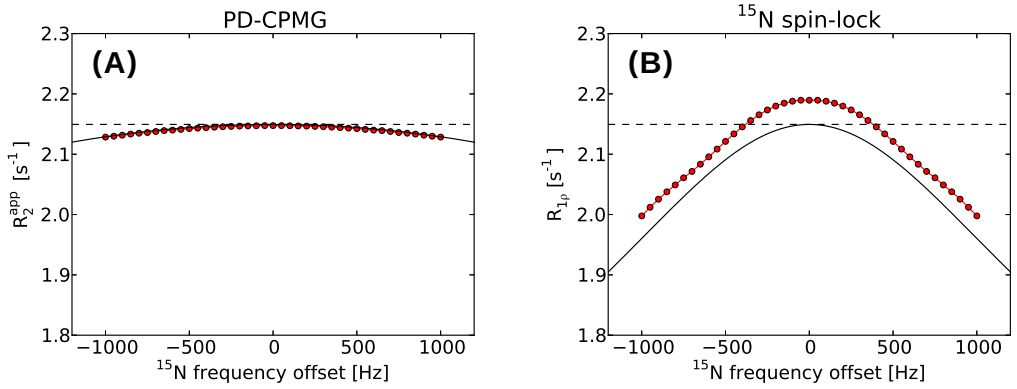


Figure 3.8. Simulated dependencies (A) R_2^{app} vs. ν_{off}^N for the PD-CPMG experiment and (B) $R_{1\rho}$ vs. ν_{off}^N for the spin-lock experiment. The results from numerical simulations are shown with red circles (red lines), from analytical formulas are shown with solid black lines, and the target value of R_2 is shown with the horizontal dashed line. The CPMG and spin-lock pulse sequence elements are modeled with the same settings as used experimentally. In particular, proton DIPSI-2 decoupling is assumed to be applied on-resonance with the *rf* field strength 4 kHz.

It can also be noted that the offset dependence illustrated in Figure 3.8A is rather weak. This is not surprising since PD-CPMG employs a train of hard ^{15}N 180° pulses which effectively refocus nitrogen magnetization across the entire spectrum. In fact, if the offset dependence is ignored the resulting error in determination of R_2 does not exceed 1%. In most cases the error of this magnitude can be tolerated and therefore it is safe to assume that $R_2 = R_2^{app}$. This obviates the need for ^{15}N R_1 measurement which is otherwise needed to calculate the offset correction according to Equation (3.1). Finally, it is worth noting that the result shown in Figure 3.8A is independent of the *rf* field inhomogeneity on both ^{15}N and ^{1}H channels (assuming any reasonable *rf* field distribution).

In Figure 3.8B it shows the ^{15}N offset dependence for ^{15}N spin-lock experiment, $R_{1\rho}$ vs. ν_N^{off}. Notably, there is a shift between the simulated $R_{1\rho}$ rates (red circles) and the standard correction formula, $R_{1\rho} = R_2 \sin^2\theta + R_1 \cos^2\theta$ (black curve). The shift has a magnitude of 0.03–0.04 s^{-1}, which mainly arises from the interplay between

the spin-lock, $^{1}J_{NH}$ coupling, and ^{1}H 180° pulses used to suppress DD/CSA cross-correlations[135]. Even though ^{1}H 180° pulses are applied infrequently, they can still give rise to a discernible error. In relative terms, this error can be safely neglected for slowly tumbling globular proteins, but not necessarily so for highly mobile IDPs.

3.1.4.2 ^{1}H Decoupling and Choice of τ_{cp}

In order to obtain accurate ^{15}N R_2 measurement from PD-CPMG, it is necessary to ensure that proton decoupling is synchronized with ^{15}N CPMG pulse train according to the synchronization condition:

$$2\tau_{cp} + p_{N(180)} = m\Delta_{DIPSI2} \tag{3.2}$$

For this purpose, one should calibrate the proton 90° pulse at the chosen decoupling power (3–4 kHz), calculate the duration of the decoupling cycle Δ_{DIPSI2}, and then set τ_{cp} accordingly.

Some undesirable properties may show up if such synchronization condition is violated. DIPSI-2 sequence is designed to achieve near-complete suppression of $^{1}J_{NH}$ over the time period Δ_{DIPSI2}. However, if at some point during Δ_{DIPSI2} one applies a ^{15}N 180° pulse this would interfere with the decoupling scheme, leading to less-than-perfect suppression of $^{1}J_{NH}$. As a consequence, nitrogen spin would experience a certain amount of scalar-coupled evolution and suffer from additional magnetization loss (e.g., due to dephasing of $2N_yH_z$ under the effect of the proton *rf* field). From the perspective of PD-CPMG measurements, this would result in overestimated ^{15}N R_2 rates. Indeed, this effect has been confirmed via simulations and experimentally—when PD-CPMG experiment is set up with no regard for synchronization conditions, the determined ^{15}N R_2 rates prove to be systematically higher which could lead to significant measurement errors.

The choice m = 2 corresponds to the strength of the effective *rf* field $\nu_1^{CPMG} = 1/4\tau_{cp} = 36$ Hz. This is much lower than the strength of the *rf* field used for proton

decoupling, 4.1 kHz. Thus, the experiment should be safe with regard to a potential risk of Hartmann–Hahn transfer. Indeed, our simulations predict that the outcome of the experiment is independent of the choice of m, confirming that Hartmann–Hahn transfer is not an issue. In principle, any integer value of m can be chosen for PD-CPMG, however, in practice too large m leads to too long τ_{cp} according to Equation (3.2), therefore relatively few points can be sampled within a given upper limit of relaxation time period. On the other hand, in reality ^{1}H decoupling is not as perfect as modeled in the numerical simulation, therefore it still relies on ^{15}N 180° pulses to aid ^{1}H decoupling and suppress $^{1}J_{NH}$. Based on these considerations, it is recommended to set ^{15}N 180° pulse interval $2\tau_{cp} \sim$ 10–15 ms, which typically corresponds to m = 1–2 for $\sim$ 3–4 kHz ^{1}H decoupling *rf* power.

Several different ^{1}H decoupling sequences have been tested for PD-CPMG. DIPSI-2 was selected primarily for its excellent J-coupling suppression properties. The decoupling bandwidth is a secondary consideration (particularly for disordered proteins with their narrow range of $^{1}H^{N}$ chemical shifts). The PD-CPMG measurements using DIPSI-3 decoupling produce the results that are virtually identical to those obtained with DIPSI-2. Choosing between the two, DIPSI-2 is preferrede since it offers a benefit of a shorter cycle. Accurate results have also been obtained using WALTZ-16. However, in this case the experiment becomes more sensitive to synchronization condition—away from the correct setting expressed by Equation (3.2) the measured R_2 rates turn out to be significantly overestimated (confirmed with both numerical simulation and experiments). Along these lines the appearance of bias in the earlier implementations of PD-CPMG can be explained[317]. Finally, GARP-1 decoupling sequence shows relatively poor performance, which is not surprising since GARP offers excellent bandwidth, but mediocre J-coupling suppression properties[344].

3.1.4.3 The Importance of Using External ^{2}H lock

It is clearly desirable to study IDPs at or near physiological conditions. However, such studies have been rare so far[345]. The main difficulties arise from fast solvent

exchange, $k_{ex} \sim 10$–$100\ s^{-1}$, which undercuts the sensitivity of HSQC experiment and worsens the situation with spectral overlaps. There are also other, more technical, problems associated with solvent exchange that can compromise the accuracy of ^{15}N relaxation measurements.

Most biological NMR samples are prepared with 5–10% D_2O for the purpose of frequency locking. However, under fast solvent exchange conditions the presence of D_2O in the sample could have a number of undesirable consequences for ^{15}N R_2 measurements[305]: (i) Standard relaxation experiments use $^1H^N(^{15}N)$ sites as a point of origin. During the course of the sequence, and in particular during τ_{rel}, the original $^1H^N(^{15}N)$ groups are partially converted into unobservable $^2H^N(^{15}N)$. This transition process modulates the recorded decay profile and thus interferes with R_2 determination. (ii) If solvent exchange is fast on the time scale of τ_{rel}, the measured R_2 rate represents a weighted average of $^{15}N(^1H^N)$ and $^{15}N(^2H^N)$ relaxation rates. (iii) In the latter case, $^{15}N(^2H^N)$ relaxation contains a sizable contribution from the scalar mechanism due to fast deuterium relaxation and solvent exchange. In PD-CPMG experiment this contribution can be only partially suppressed by infrequent application of 180° ^{15}N pulses. (iv) In addition, the apparent R_2 may include an R_{ex}-type term which stems from the isotopic frequency shift between $^{15}N(^1H^N)$ and $^{15}N(^2H^N)$.

To avoid all of these complications, it is necessary to use the external frequency lock, i.e., 100 μL D_2O placed in a coaxial insert fitted into the standard NMR tube. Such strategy has been widely used in many studies involved with fast solvent exchange labile protons[140,346], besides, it has been shown that the samples using coaxial inserts are less likely to suffer from convection flow effects[347].

3.1.4.4 Manipulating Water Magnetization

As discussed above, PD-CPMG sequence is intended for use with intrinsically disordered proteins at physiological temperature. If sample pH is also close to physiological, then solvent exchange becomes a major factor. Specifically, the accuracy of the measurements can be compromised by the transfer of the saturation from wa-

ter to amides mediated by fast exchange. Other transfer mechanisms, such as (i) intermolecular NOE and (ii) solvent exchange involving hydroxyl and amine groups followed by intramolecular NOE, can also potentially add to the problem. Additional damage can be caused by radiation damping.

Instead, we have chosen to apply water purge element[321] following the acquisition period and prior to the recycling delay. This method guarantees that the amount of water magnetization present at the beginning of each scan always remains the same. Consider the situation where solvent exchange is fast on the time scale of spin–lattice relaxation, $k_{ex} \sim 10\ s^{-1}$ or higher. In this case the recovery of amide magnetization is controlled by the water R_1 rate, which is $\sim 0.2\ s^{-1}$ at 37 °C[348]. Under these circumstances, the optimal recycling delay is approximately 5.0 s. The long recovery time means that the sensitivity of the experiment would suffer—it is $\sim$ 2-fold lower compared to the schemes that preserve water magnetization. However, this is the sacrifice that is necessary to ensure the accuracy of R_2 measurements.

3.1.4.5 Contribution from Solvent Exchange to R_2^{app}

Two potential complications associated with solvent exchange have been discussed above: (i) proton–deuterium exchange in a D_2O-containing sample and (ii) saturation transfer from H_2O to amide protons. In this section we focus on a third mechanism whereupon solvent exchange can interfere with ^{15}N R_2 measurements. In the absence of proton decoupling, the in-phase N_x magnetization interconverts with the anti-phase $2N_yH_z$. The latter is destroyed by solvent exchange, which physically separates the two correlated spins[327]. When solvent exchange is sufficiently fast on the time scale of $^1J_{NH}$, the resulting loss of magnetization can be described as scalar relaxation of the second kind[349].

In the PD-CPMG and spin-lock experiments, although $^1J_{NH}$ evolution during τ_{rel} is suppressed. the scalar relaxation is not completely absent. For spin-lock experiment, the standard expression for scalar relaxation rate $R_2^{sc} = (\pi J_{HN})^2 \tau_{ex}/(1 + (2\pi\nu_1\tau_{ex})^2)$ immediately suggests that the effect can be significant[349]. A similar outcome can be

expected for the PD-CPMG experiment, although no analytical formula is available to address the situation involving DIPSI-2 decoupling.

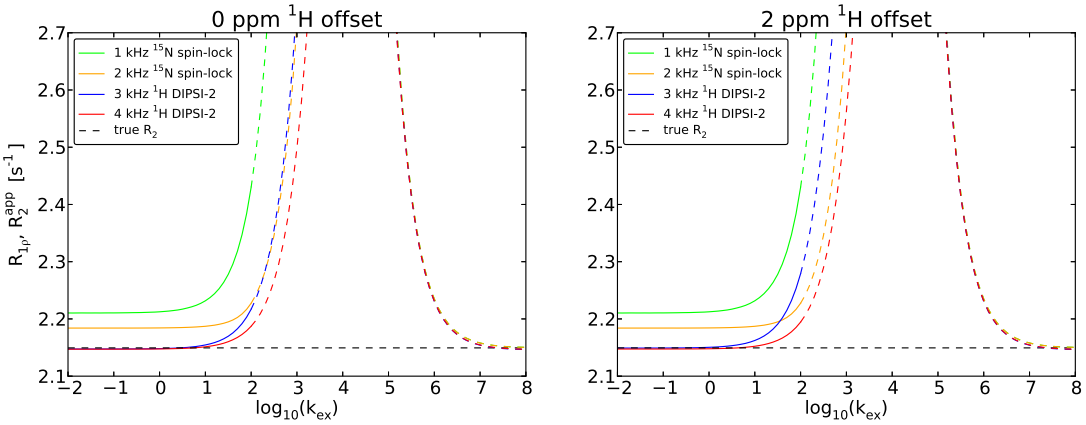


Figure 3.9. Dependence of $R_{1\rho}$ and R_2^{app} rates from ^{15}N spin-lock and PD-CPMG experiments on solvent exchange rate k_{ex}. The dependence arises from the loss of $^1H^N$–^{15}N spin correlations due to amide solvent exchange. Solid lines correspond to the conditions of practical interest; black dashed horizontal line represents the true value of R_2. k_{ex} is given in the units of s^{-1}.

The simulated $R_{1\rho}$ and R_2^{app} constants for a range of solvent exchange rates k_{ex} are shown in Figure 3.9. The graph demonstrates that slow exchange, $k_{ex} < 10$ s^{-1}, has virtually no effect on the extracted relaxation rates. As the exchange rate increases to 10–100 s^{-1}, a significant amount of error is generated. The $R_{1\rho}$ experiment using 1 kHz spin lock is particularly vulnerable (green curve). The $R_{1\rho}$ experiment using 2 kHz spin lock and PD-CPMG experiment using 4 kHz DIPSI-2 decoupling produce a similar amount of error, which is not necessarily negligible (orange and red curves, respectively). One should be mindful of this effect when conducting measurements under somewhat extreme conditions where k_{ex} exceeds several tens of s^{-1}. In this situation it may be advisable to collect the data using higher decoupling power, which helps to eliminate the contribution from solvent exchange into R_2^{app}.

The range $k_{ex} \sim 10$–$100\ s^{-1}$ is relevant for IDPs under physiological conditions[350]. This is also the range where ^{15}N relaxation can be measured using the standard HSQC-based schemes. The remaining portion of Figure 3.9 shows that the scalar contribution into $R_{1\rho}$ or R_2^{app} reaches several s^{-1} when exchange rate increases to $\sim 10000\ s^{-1}$. With further increase in k_{ex} the scalar relaxation declines. This phenomenon is known as self-decoupling; in principle, self-decoupling leads to complete suppression of $^1J_{NH}$ evolution and allows for extraction of accurate R_2 rates (all profiles in the plot converge to the dashed horizontal line). For spin-lock experiment, the numeric results shown in Figure 3.9 are in quantitative agreement with the calculations using Redfield-theory formula for scalar relaxation[349].

3.1.5 Discussion and Conclusion

The new proton-decoupled CPMG (PD-CPMG) experiment has been developed with the goal to improve the precision of ^{15}N R_2 measurements in intrinsically disordered proteins at or near physiological conditions. In this work we have demonstrated the favorable properties of this pulse sequence with denatured ubiquitin at 37 °C, 8 M urea and pH 3.0. Under real physiological conditions for studying IDPs (e.g., 37 °C, pH 7.4 and without urea), amide solvent exchange presents a potentially grave problem, and the exchange rates in disordered proteins mostly fall in the range 20–200 s^{-1}, causing significant line broadening and potentially compromising the accuracy of relaxation measurements[350]. Although PD-CPMG measurements under these conditions should be possible, it should still be extremely demanding due to too fast solvent exchange rates. In this situation it may be advisable to lower pH to 7.0, as this value remains relevant for cytosol[351]. Given that histidine side chains in IDPs titrate near pH 6.0, it should be possible to lower the pH without significantly altering the charges on the protein[352].

The ion content of the cytosol is typically assumed to be equivalent to 150 mM NaCl. We executed the PD-CPMG experiment on a sample containing this amount of salt and found that the temperature of the sample increased by 1.3 °C. This amount

of heating is tolerable and does not affect the appearance of the spectra as verified experimentally. The increase in temperature can be compensated by adjusting the setting of the VT unit. Therefore PD-CPMG measurements can be equally well conducted using the samples with physiologically relevant salt concentrations.

Finally, considering solvent viscosity and the effect of co-solvents, since the inside of the cells is known to be crowded, it may cause substantial increases in ^{15}N R_2 rates[353]. Similarly, the 8 M urea in the presence of ubiquitin measurements led to elevated R_2 values, both through increased solvent viscosity and coordination of urea to the unfolded protein[354]. In contrast, NMR samples of IDPs are usually prepared without crowding agents or osmolytes. The ^{15}N R_2 rates in such samples are typically found to be 2–3 s^{-1} at the temperatures 15–20 °C[127,355,356], which corresponds to R_2 of ~ 1 s^{-1} at the temperature of 37 °C. For such slow relaxation decay, the advantages of the proposed PD-CPMG scheme should be even more significant. In other words, the gains in precision would surpass the factor of 3 found in this study.

The issue of precision can be critical for ^{15}N relaxation measurements in IDPs. Because of the propensity to aggregate and precipitate, the samples of IDPs often need to be prepared with low protein concentration, in the range of 10–100 μM. Furthermore, the measurement time is often limited to 1–10 h because the samples tend to quickly deteriorate. Solvent exchange further compounds to the problem, causing line-broadening in the already crowded spectra and further lowering signal-to-noise ratio. Under such circumstances it can be difficult to obtain high-quality data. At the same time it is very important that the data are precise. Indeed, ^{15}N relaxation measurements in IDPs typically pursue the effects that are subtle in nature. For instance, this may be a tenuous α-helical propensity which distinguishes physiologically relevant IDP from a random-coil-like peptide chain or a putative local cluster which is formed in response to a certain point mutation. In each case we rely on small changes in ^{15}N R_2 rates to pinpoint the effect of interest therefore the data should be collected with high precision, and the new PD-CPMG experiment is expected to be quite helpful in this regard.

The accuracy of the new experiment has been established through comparison with well-established $R_{1\rho}$ scheme. Such comparison was carried out for both unfolded and folded protein samples. We have determined that PD-CPMG experiment has a number of favorable properties. First, it records pure in-phase evolution during t_1 period and therefore offers an advantage in sensitivity and resolution. The signal-to-noise ratio in the PD-CPMG spectra of α-spc SH3 and ubiquitin are $\sim 15\%$ higher than in the comparable $R_{1\rho}$ spectra. This advantage becomes dramatic for proteins or peptides experiencing significant solvent exchange. Second, PD-CPMG data can be easily corrected for ^{15}N offset effect. In fact, this correction is small and can even be neglected, thus obviating the need for companion R_1 experiment. Third, the experiment is designed in such a way that solvent exchange does not interfere with the measurements. We avoid adding D_2O to the sample and maintain the constant degree of water saturation. The contribution from scalar relaxation of the second kind arising from solvent exchange remains minimal. Fourth, the experiment is highly stable with regard to pulse miscalibration and *rf* field inhomogeneity. The amount of sample heating is on par with the $R_{1\rho}$ measurements. Fifth, possible R_{ex} contributions into ^{15}N R_2 rates, which may arise from transient oligomerization or other potentially interesting effects, are largely retained in the PD-CPMG experiment because of the low repetition rates of CPMG pulses. This is arguably preferable to the standard measurement schemes where the application of strong spin-lock field or frequent ^{15}N 180° pulses leads to partial suppression of R_{ex}. Sixth, PD-CPMG experiment has been adapted to probe N^{ε} spins in Arg side chains which are likely to be a valuable target for future studies of protein internal dynamics. All of these attractive properties make PD-CPMG ^{15}N R_2 experiment a valuable addition to the arsenal of NMR experiments targeted at intrinsically disordered proteins.

3.2 ^{1}H–^{15}N HSQC Study of Disordered Proteins with CP-HISQC

3.2.1 Introduction

Two-dimensional ^{1}H–^{15}N correlation spectroscopy is the most widely used class of experiments in the field of protein NMR. There are a number of reasons for continued popularity of these experiments. Bacterially expressed recombinant proteins with ^{15}N labeling can be produced using conventional protocols at very little extra cost. HSQC-type experiments generally have high sensitivity and yield the spectra that are well dispersed and free from interference of the water signal. Furthermore, these spectra are exquisitely sensitive to changes in protein status, such as ligand binding or conformational rearrangements. The HSQC-based sequences are also used to measure ^{15}N relaxation rates, paramagnetic relaxation enhancements (PREs), residual dipolar couplings (RDCs), and other experimental parameters of prime importance.

For all of its strengths, ^{1}H–^{15}N correlation spectroscopy is not ideally suited for measurements at or near physiological conditions (pH 7.4, 37 °C). Under these conditions the quality of the spectra deteriorates due to rapid ^{1}H^{N} solvent exchange which causes both loss of intensity and line broadening. In the case of folded soluble proteins only those amides that lie on the surface of the protein and have a significant degree of solvent exposure are affected. However, among these affected residues one often finds key catalytic residues, as well as residues which are critically important for ligand binding[357–359]. A similar situation is encountered in the samples of membrane proteins reconstituted in lipids or detergents; in this case solvent exchange has an effect on water-exposed loops[360]. Of particular importance for this study, solvent exchange can have severe adverse impact on ^{1}H–^{15}N spectroscopy of intrinsically disordered proteins (IDPs). The degree of water exposure is typically very high for most residues in IDPs. As a consequence, event at room temperature and pH 6.0 the quality of the spectra already starts to visibly deteriorate. This happens regardless of residual secondary structure. Indeed, the presence of transient, low-populated α-helices or β-turns offers little protection against solvent exchange[111,361,362]. Even

in the case of molten globule proteins, solvent exchange can thoroughly degrade the quality of ^{1}H–^{15}N spectra[363].

In principle, the problems associated with solvent exchange can be avoided if one resorts to pulse sequences employing non-labile protons or heteronuclei. A number of experiments have been developed along these lines including, notably, ^{13}C-based experiments[364–367]. However, given the substantial cost of ^{13}C-labeling and low sensitivity of the proton-less pulse schemes, the ^{1}H–^{15}N spectroscopy still remains the first choice for NMR studies of IDPs.

It has been recognized long time ago that in the presence of moderately fast solvent exchange it is important to preserve water magnetization[337,368]. The FHSQC pulse sequence is developed along these lines, which offers excellent water preservation properties[369]. Transfer of magnetization from water to labile protein sites makes it possible to relatively quickly replenish $^{1}H^{N}$ magnetization. As a result, FHSQC experiment can be executed with relatively short recycling delays. The same is true for the more recent SOFAST-HMQC experiment[370]. The superior performance of SOFAST-HMQC under moderately fast exchange conditions, $k_{ex} \sim 1$–$10\ s^{-1}$ has also been demonstrated[371].

While the experiments such as FHSQC and SOFAST-HMQC offer superior sensitivity, they also suffer from one potential setback. Specifically, these sequences utilize multispin modes such as $2N_xH_z$ or $2N_xH_x$ to record the evolution in t_1 domain. Under the effect of solvent exchange, such multispin modes undergo rapid decay, resulting in dramatic line broadening[327]. As a result, the increased solvent exchange leads to loss of spectral resolution in nitrogen dimension.

This problem has been addressed in another pulse sequence, which has been originally termed proton-decoupled HSQC[334]. This experiment relies on N_x coherence to record the evolution in t_1 domain. In the presence of robust proton decoupling, N_x evolves essentially independently of solvent exchange. In the case of highly mobile disordered protons, the spectra can be recorded with very narrow linewidth in ^{15}N dimension. The favorable properties of proton-decoupled HSQC with respect to sol-

vent exchange have been pointed out already in the original report[334]. Later the same strategy has been used to record the spectra of lysine side-chain NH_3 groups, which are otherwise difficult to observe due to the fast solvent exchange[140]. The lysine experiment was termed heteronuclear in-phase single quantum coherence spectroscopy (HISQC). This scheme has also been used in the context of ^{15}N R_2 measurement for IDPs in PD-CPMG pulse sequence[132].

In this report we present the proton-decoupled HSQC experiment which uses cross-polarization (CP) element as a starting point. The new pulse sequence is termed CP-HISQC. As demonstrated below, the CP element makes efficient use of water magnetization, achieving the following magnetization transfer:

$$H_x^{water} \xrightarrow{k_{ex}} H_x^N \xrightarrow{^1J_{NH}} N_x \tag{3.3}$$

For the sample conditions that are close to physiological, the use of the CP scheme affords significant gain in sensitivity compared to the previously reported HISQC sequence. At the same time, CP-HISQC offers dramatically improved spectral resolution compared to conventional experiments such as FHSQC and SOFAST-HMQC. The combination of superior resolution and good sensitivity makes CP-HISQC sequence particularly well-suited for studies of IDPs at or near physiological conditions. It can also be useful for studies of globular proteins or membrane proteins if the goal is to capture the signals from surface sites undergoing rapid solvent exchange.

Setting up the experiment is sufficiently easy, and the efficient CP transfer is achieved simultaneously for those residues that experience fast solvent exchange and for those residues that are protected. This property has been demonstrated on the sample of N-terminal SH3 domain of adaptor protein drk, which contains both folded and unfolded forms of the protein[372]. Using CP-HISQC sequence we have recorded high-quality backbone $^1H^N$–^{15}N spectra for drkN SH3 at 30 °C, pH 7.5. In addition, we have also recorded arginine side-chain $^1H^{\varepsilon}$–$^{15}N^{\varepsilon}$ spectra of ten-residue peptide from guanine nucleotide exchange factor Sos[197]. These data were collected at pH 6.0

in the range of temperatures from 23 to 37 °C. Using the sample of Sos peptide we have also shown that CP-based $^{1}H^{N} \rightarrow {}^{15}N$ magnetization transfer step still remains efficient even in the presence of very fast solvent exchange under the condition of 50 °C, $k_{ex} \sim 600\ s^{-1}$.

3.2.2 Materials and Methods

Sos peptide has been expressed in ^{13}C, ^{15}N enriched M9 media in a form of GB1–Sos fusion construct and purified by cation-exchange and gel-filtration FPLC. The peptide was subsequently cleaved by application of formic acid[373] and purified by reverse phase HPLC. The concentration of the peptide has been determined by comparing the volume of V4 H^{γ} peak in 1D ^{1}H spectrum of Sos with the corresponding signal from 10 mM reference sample of valine. The conditions of the Sos sample were 1.0 mM Sos, 20 mM phosphate buffer in H_2O, pH 6.0, 0.02% NaN_3. drkN SH3 has been expressed in ^{15}N enriched M9 media and purified according to the standard protocol[374]. Protein concentration has been determined via UV absorbance at 280 nm. The sample conditions were 2.0 mM drkN SH3, 20 mM phosphate buffer in H_2O, pH 7.5, 0.02% NaN_3. In both samples, Sos and drkN SH3, D_2O has been added into coaxial insert (Wilmad, WGS-5BL) to serve as external lock. This strategy avoids numerous complications that arise from the use of H_2O/D_2O solvent under fast solvent exchange conditions[132,305].

All NMR measurements were performed on Varian Inova 600 MHz spectrometer equipped with a z-gradient HCN triple-resonance probe. 1D CP-HISQC experiment has been used to optimize *rf* field settings for the CP element. In doing so, ^{15}N power levels were scanned with fixed ^{1}H power level. The CP-HISQC arginine side-chain experiments on Sos have been recorded at 23, 30, and 37 °C (temperature calibrated using methanol). The spectra were collected with spectral width 6 ppm in ^{15}N dimension (t_1 = 175 ms) and 15 ppm in ^{1}H dimension (t_2 = 46.5, 23.3, and 11.8 ms for the three respective temperatures, chosen according to the formula $t_2 = 2/k_{ex}$ since the decay of the useful signal is dominated by solvent exchange). The data were

apodized using phase-shifted sine-bell squared window function and zero-filled to the size 512 × 4096. The CP-HISQC spectrum of drkN SH3 was recorded at 30 °C with spectral width 30 ppm in ^{15}N dimension and 15 ppm in ^{1}H dimension (t_1 = 177 ms, t_2 = 40 ms, chosen according to the formula $t_2 = 2/k_{ex}$ based on the compromise value $k_{ex} = 50\ s^{-1}$ for solvent exchange rate). The data were apodized and zero-filled to the size 2560 × 4096. The peak intensities have been obtained using nlinLS subroutine in NMRPipe[261]. The average noise level was estimated with the help of the showApod subroutine. The solvent exchange rates were measured based on previous developed strategy[375] implemented with some minor modifications (e.g., the altered sequence evolves N_x coherence during the t_1 period instead of $2N_xC_z$).

The efficiency of proton-to-nitrogen CP and INEPT transfer in two-spin ^{1}H–^{15}N system has been simulated using in-house Maple and Python scripts. All coherent evolution mechanisms such as chemical shifts, J-couplings and *rf* pulses have been included in the simulations in a fully realistic fashion. The modeling of solvent exchange effect follows the previous described scheme[327], where all two-spin modes experience decay with the rate k_{ex}, whereas H_x^N, H_y^N and H_z^N modes exchange magnetization with the corresponding water modes H_x^{water}, H_y^{water} and H_z^{water}. The ratio of water magnetization to protein proton magnetization is assumed to be 1000:1, reflecting protein concentration ∼ 1 mM and assuming that a small-sized protein contains 100 exchangeable protons. The results do not change if the ratio is assumed as 10000:1, corresponding to ∼ 0.1 mM sample. Given that spin relaxation in IDPs at or near physiological temperature is relatively slow and the mixing time τ_{mix} is short, we ignore DD and CSA spin relaxation. Likewise, we have neglected the potential effect of radiation damping noticing that water magnetization is either locked by strong *rf* field during CP element or dephased by application of pulsed field gradients during INEPT element. The size of the Liouville superoperator matrix used in the simulations is 18 × 18 (comprising 15 modes from ($^{1}H^N$,^{15}N) spin system and 3 modes from water ^{1}H spin). The initial conditions have been chosen assuming that equilibrium magnetization is available for both $^{1}H^N$ and water. Alternatively, we considered

somewhat artificial scenario where the water is initially saturated, but $^{1}H^{N}$ retains the full amount of magnetization (see Figure 3.14). More details about the numerical simulation setup for CP-HISQC are included in Appendix C.

3.2.3 Results

3.2.3.1 CP-HISQC Pulse Sequence

The new pulse sequence CP-HISQC (cross-polarization assisted heteronuclear in-phase single-quantum correlation) is shown in Figure 3.10. The initial hard ^{1}H 90° pulse followed by high-power DIPSI-2 pulse train lock both amide and water magnetization along the x axis. As amides undergo solvent exchange, the H_x^N magnetization on amides is gradually replaced with H_x^{water} magnetization from water. At the same time H_x^N is continuously transformed into N_x via the cross-polarization transfer (Equation (3.3)). The ability of the CP element to efficiently utilize the influx of new water magnetization is the most important factor for the superior performance of CP-HISQC.

Following the CP element, N_x magnetization is ^{15}N-frequency-labeled during the evolution period t_1, while proton decoupling is applied to suppress $^{1}J_{NH}$ and thus prevent the build-up of anti-phase magnetization. The use of the in-phase nitrogen magnetization greatly improves the quality of the spectra compared to the standard scheme (where $2N_xH_z$ evolves during t_1, suffering dramatic losses due to rapid solvent exchange). Following the evolution period, refocused INEPT is used to return magnetization to amide protons for detection. While CP element is strongly preferred for the initial proton-to-nitrogen transfer, there are much fewer advantages of CP over other schemes (such as INEPT) during the reverse magnetization transfer step (i.e., $^{15}N \rightarrow {}^{1}H^{N}$), therefore CP element is only employed during the initial step of coherence transfer (i.e., $^{1}H^{N} \rightarrow {}^{15}N$).

Special care has been taken to preserve water magnetization throughout the pulse sequence. As already pointed out, water magnetization is locked during the CP

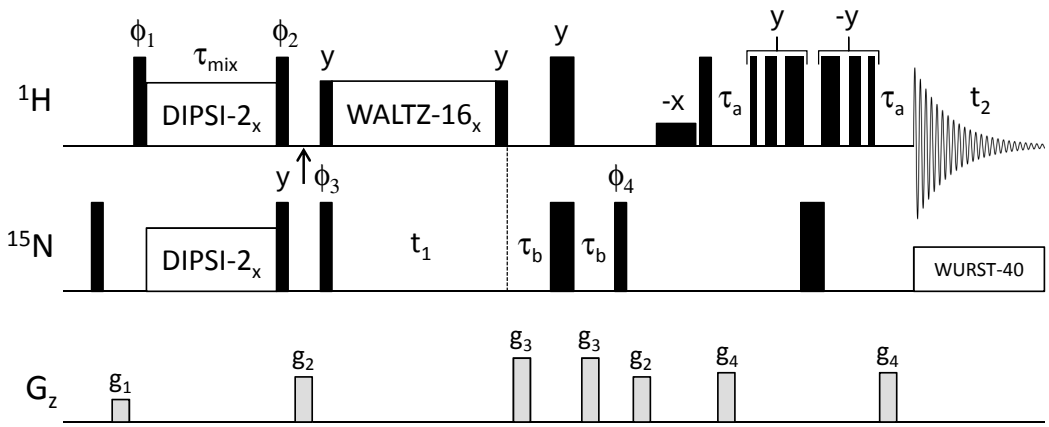


Figure 3.10. CP-HISQC pulse sequence for ^{1}H–^{15}N correlation spectroscopy. Proton carrier is initially placed at the middle of the amide $^{1}H^{N}$ region ($\sim$ 8.20 ppm). At the point indicated by vertical arrow, the carrier is jumped to water resonance. The durations of delays are $\tau_a = 0.8 \times 1/(4J_{NH}) = 2.15$ ms and $\tau_a = 0.9 \times 1/(4J_{NH}) = 2.40$ ms that have been optimized assuming $k_{ex} = 100$ s^{-1}. The spacing between the consecutive pulses in 3-9-19 WATERGATE[376] is 183 μs. The phase cycle employed is: $\phi_1 = y, -y$; $\phi_2 = -y, y$; $\phi_3 = 2(y), 2(-y)$; $\phi_4 = 4(x), 4(-x)$; $\phi_{rec} = x, -x, -x, x, -x, x, x, -x$. DIPSI-2 and WALTZ-16 sequences are applied with initial phase along x. Gradient strengths in G/cm (length in ms) are: $g_1 = 5.0$ (1.0), $g_2 = 10.0$ (1.0), $g_3 = 15.0$ (1.25), $g_4 = 11.0$ (1.0).

element; likewise, it is locked during t_1 period by means of the WALTZ sequence of phase x. Finally, during the refocused INEPT water magnetization is maintained along the z axis. As a result, at the end of each scan water retains $\sim$ 75% of its equilibrium magnetization, which is then quickly transferred to amide sites through solvent exchange. In this manner it avoids long recycling delays which would otherwise be necessary to allow for recovery of water magnetization.

The sequence shown in Figure 3.10 is similar to the well-known proton-decoupled HSQC experiment[334], as well as the more recent HISQC experiment[140] developed for Lys side-chain NH_3 groups. The main distinction of CP-HISQC pulse sequence is the use of CP element which leads to significantly better sensitivity. We have also found that 3-9-19 WATERGATE[376] offers better water preservation compared to the version that employs 90° soft rectangular pulses[140].

The CP element has been designed such as to lock both water and $^1H^N$ magnetizations along the x axis. We have determined that optimal results are obtained when (i) carrier is set on $^1H^N$ and (ii) the strength of DIPSI *rf* field is at least two times greater than the frequency offset between $^1H^N$ and water (i.e., $\nu_1^H/\Delta \geq 2.0$). In principle, increasing the *rf* field strength beyond this level leads to even better results, but the coherence transfer efficiency is only slightly higher than $\nu_1^H/\Delta = 2.0$.

Strictly speaking, the *rf* field strength $\nu_1^H = 2\Delta$ is insufficient to properly align water magnetization along the x axis. Following the initial hard 90° pulse, water magnetization lands on the x axis. Subsequently, under the effect of off-resonance DIPSI-2, water magnetization travels on a complicated trajectory on the surface of a unit sphere (while maintaining a significant x projection). This situation can be described in terms of water magnetization being partially locked. After full DIPSI-2 cycle is completed, water is returned to the x axis, reflecting excellent offset compensation properties of the DIPSI-2 composite pulse. At this point the second hard 90° pulse is applied, returning water magnetization to the z axis. This scheme ensures good preservation of water magnetization, which is essential for the sensitivity of the measurements.

The necessary condition for this favorable scenario is that the length of the CP period, τ_{mix}, is proportional to the integer number of DIPSI-2 cycles. In practice, we have found that good results are obtained already with semi-integer number of DIPSI-2 cycles, $\tau_{mix} = (n/2) \times T_{DIPSI2}$. On the other hand, the length of τ_{mix} is dictated by the requirements of cross-polarization transfer. In the absence of solvent exchange, the optimal transfer is achieved when $\tau_{mix} = 1/^1J_{NH} = 10.8$ ms. Although it seems hard to reconcile these two requirements at the same time, however, it turns out that in the presence of rapid solvent exchange, high efficiency of the CP transfer can be attained in a broad interval of τ_{mix}. Specifically, the range of $\tau_{mix} \sim$ 10–13 ms can be recommended (see Figure 3.14). Within this range it is easy to accommodate

τ_{mix} as semi-integer number of DIPSI-2 cycles. For example, for the study on drkN SH3 backbone amide group, $\tau_{mix} = (3/2) \times T_{DIPSI2} = 10.705$ ms was used.

3.2.3.2 CP-HISQC Spectra of $^{15}N^{\varepsilon}$–$^{1}H^{\varepsilon}$ Groups from Arg Side Chains

Sos peptide with the primary sequence PPPVPPRRRR is a proline-rich fragment from guanine nucleotide exchange factor Sos. This fragment is responsible for the interaction with SH3 domains of various adaptor proteins[197]. The signals from four arginine $^{15}N^{\varepsilon}$–$^{1}H^{\varepsilon}$ groups in Sos are largely overlapped; the spectrum contains two resolved peaks of unequal height (see Figure 3.11, bottom row).

Solvent exchange rates for $^{15}N^{\varepsilon}$–$^{1}H^{\varepsilon}$ sites can be measured by means of the experiment that records ($^{13}C^{\delta}$,$^{1}H^{\delta}$) correlation map[375]. In the case of Sos peptide, this spectrum features a single four-fold degenerate peak. As a consequence, k_{ex} constant obtained in this experiment represents an effective average of the four exchange rates. Given that four arginines are contiguously located at the C-terminus of the Sos peptide and are dynamically disordered, it is reasonable to assume that their individual exchange rates are all similar to each other and close to the experimentally measured k_{ex}. Using ^{13}C-labeled sample of Sos with pH 6.0 and pulse sequence adapted from previous developed scheme[375], we determined $^{15}N^{\varepsilon}$–$^{1}H^{\varepsilon}$ solvent exchange rate at three different temperatures: 43 s^{-1} at 23 °C, 86 s^{-1} at 30 °C and 170 s^{-1} at 37 °C. This range of k_{ex} is well suited for testing the performance of the new CP-HISQC experiment.

The $^{15}N^{\varepsilon}$–$^{1}H^{\varepsilon}$ spectral maps from the new CP-HISQC experiment have been compared with the spectra from four other popular experiments with optimized sensitivity and/or resolution: SOFAST-HMQC[108], FHSQC[369], HISQC[140], and HISQC-wg3919 (the modification for original HISQC sequence that utilizes 3-9-19 WATERGATE element[376]). The latter three experiments are specifically designed for samples with fast solvent exchange. All spectra were recorded using the recycling delays optimized for best signal-to-noise ratio: 0.2 s for SOFAST-HMQC and FHSQC, 1.0 s for HISQC, HISQC-wg3919, and CP-HISQC. When choosing the recycling delay for CP-HISQC,

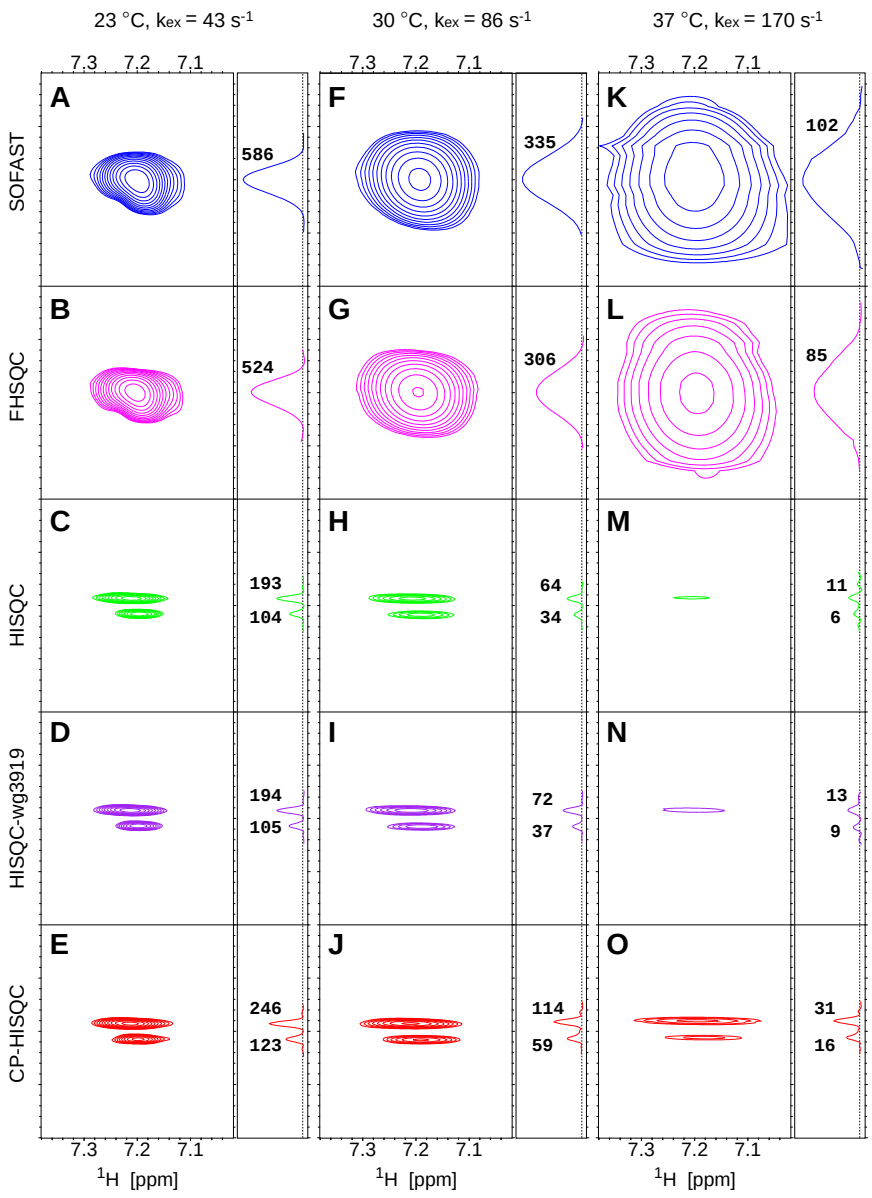


Figure 3.11. Arginine $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ correlation maps obtained for the sample of Sos peptide at three different temperatures using five different experimental schemes (as indicated in the plot). The side panels show 1D cross-sections at 7.2 ppm; the number next to each peak indicates the corresponding signal-to-noise ratio. All spectra measured at the same temperature are plotted with the same scale. The separation between the two $^{15}N^{\varepsilon}$ resonances is ~ 0.3 ppm.

one should bear in mind that cross-polarization element creates a certain load on the probe (4.0 kHz ^{15}N *rf* field applied for 10.7 ms), although this is not a factor for the latest-generation probes which can easily handle such amount of power and require only short recycling delays[315]. All experiments illustrated in Figure 3.11 use the same INEPT delays, which have been optimized based on model calculations assuming k_{ex} = 100 s^{-1}.

The data in Figure 3.11 show that the SOFAST-HMQC and FHSQC sequences which rely on the evolution of two-spin coherences ($2N_xH_x$ and $2N_xH_z$, respectively) suffer from fast solvent exchange, causing broadening of spectral peaks in indirect dimension. In contrast, the three HISQC sequences which record the evolution of in-phase coherence (N_x) produce much sharper signals. Given this distinctive behavior, we have set up the experiments such as to obtain optimal performance in relation to their intrinsic properties.

For simplicity let us focus on the data in the right column, where the effect of fast solvent exchange is especially prominent. The SOFAST-HMQC and FHSQC data shown in the right column have been recorded with only 4 increments in indirect dimension (t_1 = 8 ms); because of the rapid loss of magnetization there is no need in lengthy evolution period. The line broadening is extreme and the spectral resolution is poor, making it impossible to identify two distinct peaks separated by $\sim$ 0.3 ppm in ^{15}N dimension. Due to excellent water preservation properties, $^1H^\varepsilon$ magnetization is quickly replenished in these experiments. As a consequence, the data can be recorded with high repetition rate, which makes the experiments highly sensitive (cf. panels K and L in Figure 3.11).

On the other hand, all three HISQC sequences have been recorded with 64 increments (t_1 = 175 ms). The decay of N_x magnetization is slow, which allows for long t_1 evolution and translates into high resolution in ^{15}N dimension. If desired, the resolution can be further improved by extending t_1 period, thus taking advantage of the fact that intrinsic ^{15}N R_2 relaxation in peptides and disordered proteins is typically very slow at the physiological temperature of 37 °C, on the order of $\sim$ 1 s^{-1}. In

principle, it is possible to obtain spectra with very sharp ^{15}N resonances along these lines, reminiscent of small molecules. Recording such spectra would require somewhat longer experimental time, $\sim$ 1–2 h.

While HISQC experiments clearly display superior resolution, their sensitivity is a source of concern. A substantial amount of signal is lost due to solvent exchange during the refocused INEPT periods. Furthermore, the water preservation properties are not as good as in SOFAST experiment, which necessitates the use of longer recycling delays. All of this translates into mediocre sensitivity. The panel M in Figure 3.11 illustrates arginine $^{15}N^{\varepsilon}$–$^{1}H^{\varepsilon}$ spectrum recorded by means of the HISQC sequence[140]. The two peaks are borderline detectable, with S/N ratios of 11 and 6. Certain improvement can be obtained by implementing a better water preservation scheme. Specifically, we have found that 3-9-19 WATERGATE[376] achieves better results than the original version employing water flip-back pulses. To quantitate this (trivial) improvement, we have set up the experiment which is different from the original HISQC only in that it uses 3-9-19 WATERGATE (HISQC-wg3919). The obtained moderate improvement can be judged by examining the data in panel N.

Of greater interest, however, are the gains associated with the use of CP element to transfer magnetization from ^{1}H to ^{15}N. Replacing the first refocused INEPT period in the pulse sequence with CP element results in twofold gain in peak intensity (cf. panels N and O in Figure 3.11). The S/N ratios of the two observed peaks are increased to 31 and 16. Of note, CP-HISQC is the only experiment which obtains a fully satisfactory spectrum in $\sim$ 10 min, allowing for reliable detection of the two well-resolved peaks.

Finally it is worth pointing out that original HISQC sequence has been developed for lysine side-chain NH_3 groups; in these four-spin systems the maximum attainable efficiency for both refocused INEPT and CP element does not exceed $\sim$ 44%, however, CP-HISQC still retains its advantage over the regular HISQC also in this case. On

the other hand, the application to arginine side-chain $H^{\varepsilon}N^{\varepsilon}$ or backbone amide $H^{N}N$ sites does not suffer from such efficiency loss due to the property of two-spin systems.

3.2.3.3 CP-HISQC Spectra of Backbone Amides

The N-terminal SH3 domain of the *Drosophila* adaptor protein drk (drkN SH3 domain) is a well-established model system for studies of protein disorder[377]. The protein is marginally stable; *in vitro* it exists in a state of dynamic equilibrium between folded (F) and unfolded (U) forms. The exchange between the two states is slow, so that the spectra feature two distinct sets of peaks, therefore it is an attractive model for complex protein system, which combines the elements of order and disorder. Such complex systems increasingly become the focus of advanced NMR studies[378].

For our experimental measurements we have used the sample of drkN SH3 at pH 7.5, 30 °C. The proportion of F and U species under these conditions is approximately 1:1[374]. The unfolded form of drkN SH3 offers little protection against solvent exchange[265]. Therefore, it is appropriate to calculate solvent exchange rates using the simple algorithm which has been originally developed for random-coil peptides[350,379]. The results are dependent on protein primary sequence; for the unfolded drkN SH3, the majority of the predicted k_{ex} values fall in the range 15–150 s^{-1}, which just belongs to the regime where CP-HISQC is expected to be useful.

In what follows we compare CP-HISQC with four other experiments that have been designed with fast solvent exchange in mind. Similar to the previous section, we have set up each experiment with consideration for its individual properties. Specifically, SOFAST-HMQC and FHSQC have been recorded with 72 increments in indirect dimension, $t_1 = 39$ ms, and short recycling delays, $d_1 = 0.2$ s. At the same time, all HISQC experiments were recorded with 320 increments, $t_1 = 177$ ms, and standard recycling delay, $d_1 = 1.0$ s. Similar to the previous section, the best results have been obtained from SOFAST-HMQC (high sensitivity) and CP-HISQC (superior resolution in combination with reasonably good sensitivity). The corresponding spectra, both of which have been collected in ~ 50 min, are shown in Figure 3.12.

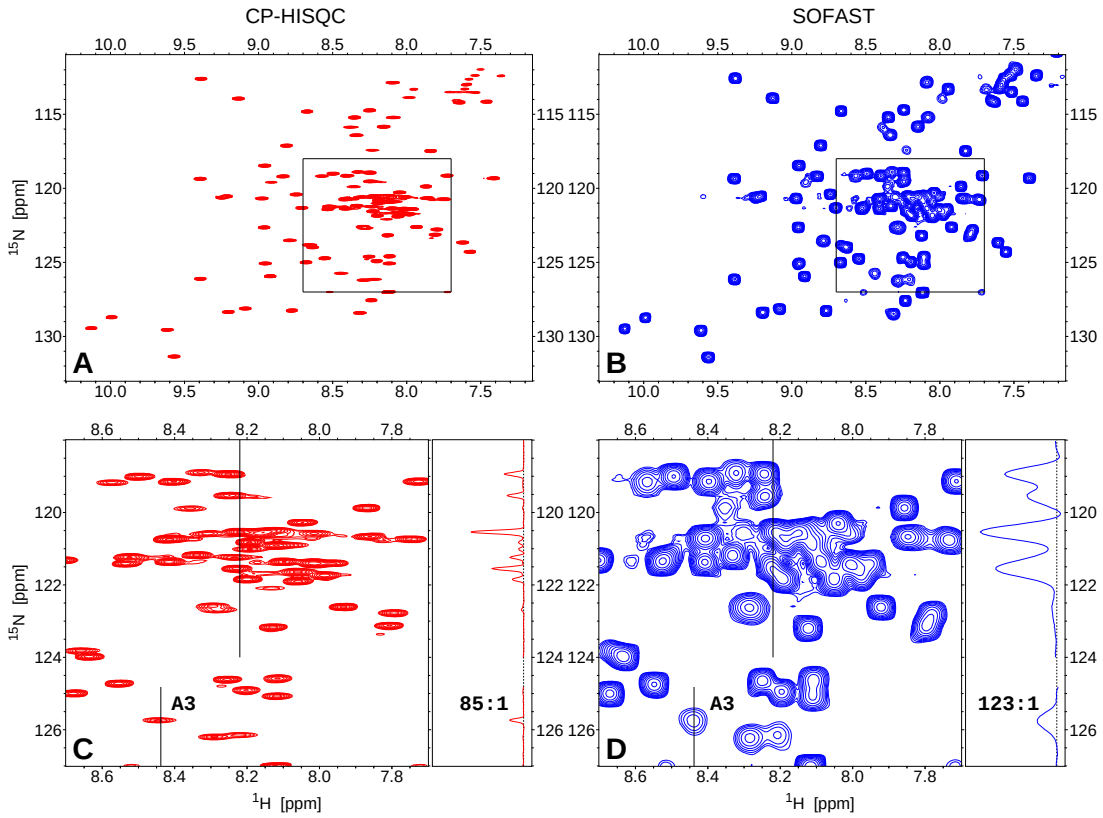


Figure 3.12. (A) CP-HISQC and (B) SOFAST-HMQC backbone amide spectra of drkN SH3. The regions enclosed in rectangular boxes are magnified in panels (C, D). Black vertical lines indicate the position of 1D traces (displayed in side panels). The peak from residue A3 belonging to the unfolded form of drkN SH3 is labeled in the graph. For the purpose of plotting, the spectra were scaled such that the noise level in all two-dimensional spectral maps is the same.

The results in Figure 3.12 demonstrate a dramatic improvement in resolution afforded by CP-HISQC experiment. The effect is especially obvious in panels C and D, which represent the most crowded portion of the spectrum. Indeed, a significant number of signals that are partially or fully overlapped in the SOFAST-HMQC spectrum turn out to be nicely resolved in CP-HISQC (cf. also 1D traces displayed in the side panels). This result is of major importance for studies of intrinsically disordered proteins or, more broadly, protein systems with elements of disorder. It is well known

that the spectra of IDPs tend to be crowded and therefore high resolution is often key to the success of NMR studies. The increase in the number of quantifiable peaks should be very useful for various HSQC-based titration experiments. It also makes CP-HISQC a potentially useful template for ^{15}N R_2 experiment, PRE experiment, and other similar measurements.

In addition to superior spectral resolution, CP-HISQC experiment also offers reasonably good sensitivity. This can be appreciated by looking at A3 resonance which is labeled in the spectral map. The signal-to-noise ratio for this peak, which belongs to the U subspectrum, is 85:1 (see Figure 3.12C, side panel). This is comparable to the S/N ratio found in the SOFAST-HMQC spectrum, 123:1 (Fig. 3.12D, side panel).

To further characterize the sensitivity of the two experiments, we have selected 22 peaks belonging to the U subspectrum which are well resolved and can be unambiguously assigned in all of the spectra[380]. For each of these spectral peaks we have determined the signal-to-noise ratio. The results are shown in Figure 3.13A, where the x axis represents the intensity of CP-HISQC signals, while the y axis represents the relative intensity of CP-HISQC with respect to SOFAST-HMQC. On average, the sensitivity of the CP-HISQC experiment proves to be 0.51 of that achieved by SOFAST-HMQC. It is worth noting, however, that CP-HISQC becomes increasingly competitive for the weaker signals (cf. the negative correlation in Figure 3.13A). For the weakest peak, that from residue S10, the sensitivity of the CP-HISQC experiment relative to SOFAST-HMQC reaches 0.78.

A similar comparative analysis can be performed for the peaks from the F subspectrum. For the most part the F peaks are insensitive to exchange and therefore have much higher intensity than U peaks. Using the data from 44 well-resolved F resonances, we have found that the relative sensitivity of CP-HISQC compared to SOFAST-HMQC on average amounts to 0.36. In the case of surface residue N35, which experiences fast solvent exchange, this parameter rises to 0.57. In this connection it should be noted that CP-HISQC experiment has been designed to target U

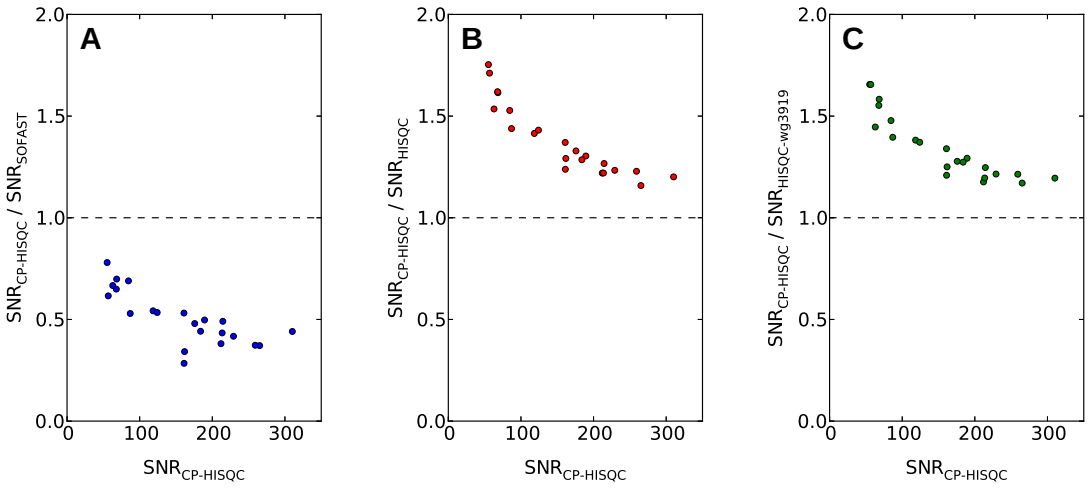


Figure 3.13. Sensitivity of CP-HISQC experiment relative to SOFAST-HMQC, regular HISQC, and HISQC-wg3919 experiments (A, B, and C, respectively). Plotted along x axis: signal-to-noise ratio (SNR) of the individual peaks in the U subspectrum of CP-HISQC spectrum of drkN SH3. Plotted along y axis: relative signal-to-noise ratio of the same peaks, as obtained from comparison with other experiments. The SNR values are calculated using peak intensities (i.e., fitted maximum amplitudes), normalized by the noise level.

sites rather than well-protected F sites. At the same time, SOFAST-HMQC is known to be very efficient for F species, i.e., small globular proteins.

Of note, standard experiments such as SOFAST-HMQC cannot be easily optimized for simultaneously collecting data from U and F components at or near physiological conditions. Indeed, in U species the $2N_xH_x$ mode which is employed in SOFAST-HMQC undergoes rapid decay due to solvent exchange, $k_{ex} \sim 15$–150 s^{-1}. On the other hand, in F species this mode is relatively long-lived, $R_{DQ/ZQ} \sim 1$–5 s^{-1} assuming the temperature 30–37 °C. This imposes conflicting requirements on the length of the t_1 evolution period. The new CP-HISQC experiment is free of this complication, since CP-HISQC utilizes the N_x mode which decays slowly in both U and F species.

Of major interest to us is the contribution of the CP element. Figure 3.13B illustrates the sensitivity of CP-HISQC experiment relative to the previously reported HISQC experiment. For those U peaks that are well resolved and unambiguously assigned, the average increase in sensitivity proves to be 1.38. The most significant gain, 1.75, is again observed for the weakest peak, S10.

In part, these gains are due to the more efficient water preservation scheme implemented in CP-HISQC. To quantify this factor, we have implemented the reference experiment HISQC-wg3919. The only difference between CP-HISQC and HISQC-wg3919 is that the former starts with CP element, while the latter relies on refocused INEPT. The relative sensitivity of the two measurements is illustrated in Figure 3.13C. The average gain from the use of the CP element is described by the factor 1.34.

Finally, CP element also works well for F species. All of the 44 selected F peaks are stronger in the CP-HISQC spectrum than in the HISQC-wg3919 spectrum. On average, the improvement is expressed by the factor 1.10. The biggest gain, 1.38, is again found in residue N35.

Based on all of the above, we suggest that CP-HISQC is the best choice of experiment for $^{1}H^{N}$–^{15}N spectroscopy of drkN SH3 at or near physiological conditions. The HISQC scheme achieves critical improvement in resolution, whereas the CP element offers significant boost in sensitivity.

3.2.4 Discussion

3.2.4.1 The Efficiency of CP Scheme: Dependence on τ_{mix}

The advantage of CP-HISQC over more conventional schemes, such as HSQC, FHSQC, SOFAST-HMQC, mainly stems from three factors:

(i) The reliance on the in-phase N_x evolution during the t_1 period. It is well appreciated that N_x has more favorable relaxation properties compared to

$2N_xH_z$ or $2N_xH_x$[334]. In the context of this study it is especially important that N_x is immune to the effects of solvent exchange, whereas $2N_xH_z$ and $2N_xH_x$ suffer heavy losses[140].

(ii) The ability of the CP element to efficiently utilize the influx of H_x magnetization from the H_2O proton pool, converting H_x into the desired N_x magnetization.

(iii) The more favorable properties of the CP element relative to the INEPT scheme with regard to spin relaxation[381] as well as solvent exchange. Both CP and refocused INEPT generate multispin modes on the route from H_x to N_x. However, the refocused INEPT has a higher content of multispin modes and, therefore, suffers greater losses due to solvent exchange.

The advantages associated with property (i) have been previously documented, although to the best of our knowledge the application to backbone amides undergoing fast solvent exchange is new. This aspect is also illustrated above (cf. the sharpening of spectral peaks in the HISQC family of spectra, Figures 3.11 and 3.12). Therefore in what follows we focus on properties (ii) and (iii) wherein lies the novelty of our approach.

To analyze the details of CP transfer in the presence of fast solvent exchange we have conducted a series of numeric simulations. Briefly, we have simulated the behavior of the two-spin system, $^1H^N$–^{15}N, undergoing exchange with the large water proton pool. For simplicity, we have neglected DD and CSA relaxation. Other aspects, such as DIPSI-2 pulse train which is applied on-resonance or close to resonance with $^1H^N$ signal but off-resonance relative to the water line, are modeled in a fully realistic fashion. In Figure 3.13 we analyze the efficiency of the CP transfer, i.e., the amount of the generated N_x magnetization, as a function of the mixing time τ_{mix}. Since the focus of this simulation is on the CP element, water preservation outside the CP element is of no interest at this point. Accordingly, we disregard the synchronization condition $\tau_{mix} = (n/2) \times T_{DIPSI2}$.

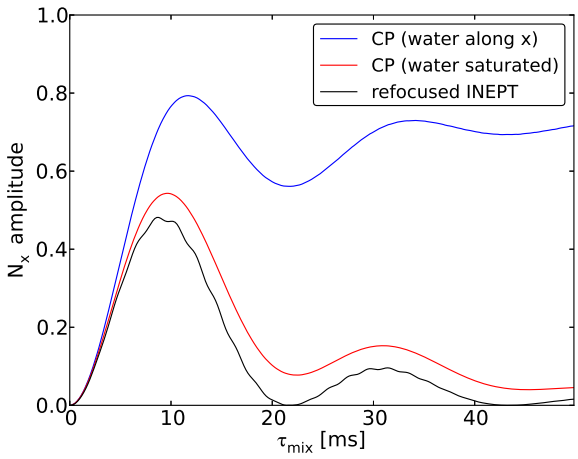


Figure 3.14. Numerical simulations of proton-to-nitrogen magnetization transfer using CP or refocused INEPT element in the presence of solvent exchange with $k_{ex} = 100\ s^{-1}$. The simulations model the conditions and setup of the actual backbone experiment, including the use of DIPSI-2 sequence with $\nu_1^H = \nu_1^N = 4.0$ kHz and proton *rf* carrier positioned on-resonance with $^1H^N$ spin at 8.2 ppm. Water magnetization is assumed to be initially aligned along x (blue and black profiles) or otherwise saturated (red profile). The refocused INEPT is assumed to be slightly asymmetric, $\tau_a/\tau_b = 0.9$, same as in the experimental measurement.

Shown in Figure 3.14 are three different profiles. The blue curve corresponds to the CP element as implemented in our experiment. Initially it is assumed that full amount of ^1H magnetization is available both on the amide and the water protons. Water ^1H magnetization is first aligned along the x axis and then effectively locked by the DIPSI-2 train. The resulting transfer profile is very different from the conventional $(1/2)(1 - \cos(\pi J_{HN}\tau_{mix}))$ oscillatory pattern normally associated with the cross-polarization transfer. As can be seen from the plot, the transfer efficiency actually converges towards the plateau value of ~ 0.70. This behavior can be understood by noticing that fresh proton magnetization is constantly injected into the amide site via solvent exchange and then transformed into N_x. After sufficiently long τ_{mix} the system reaches a state of quasi equilibrium, $H_x \rightleftarrows 2(H_yN_z + H_zN_y) \rightleftarrows N_x$, where the

loss of H_yN_z, H_zN_y due to solvent exchange is compensated by continuous influx of H_x from the large water pool.

The maximum transfer efficiency 0.79 is achieved with τ_{mix} of 11.6 ms. This is somewhat longer than the standard duration of the cross-polarization period, $\tau_{mix} = 1/^1J_{NH} = 10.8$ ms. The additional time is required to utilize a portion of the incoming water polarization, maximizing the output of N_x. Note, however, that transfer efficiency remains very good in a broad range of τ_{mix}, from ~ 10 to ~ 13 ms. This provides us with the freedom to adjust τ_{mix} such as to satisfy the requirement of $\tau_{mix} = (n/2) \times T_{DIPSI2}$. The distinctive dependence on τ_{mix} as seen in Figure 3.14 has also been verified experimentally.

The red profile in Figure 3.14 describes the situation which is in every way similar, except it is assumed that water magnetization is saturated and remains zero throughout the course of the CP period. This scenario leads to a rather different outcome. Initially, amide magnetization H_x is transformed into N_x, with maximum transfer efficiency 0.54 reached at $\tau_{mix} = 9.6$ ms. As mixing time is increased beyond this level, amide $^1H^N$ magnetization is nullified due to the exchange with saturated water magnetization. Furthermore, the existing N_x magnetization is transformed back into H_x and subsequently destroyed through solvent exchange. As a result, the efficiency of the transfer drops to a very low level.

Finally, the black profile in Figure 3.14 corresponds to the refocused INEPT transfer, where $\tau_{mix} = 2\tau_a + 2\tau_b$. In this case the behavior of water magnetization is different and deserves a special discussion. During the delays τ_a both amide and water ^1H magnetizations undergo free precession. Their precession frequencies are different by several thousand Hz. As a result, when water magnetization enters the amide site, it does so with a random phase relative to the amide magnetization. Consequently, the input from water magnetization cancels out and the situation becomes generally similar to the previous scenario, where water signal is saturated.

The main lessons that can be learned from these results is that the biggest advantage associated with CP sequence stems from its ability to utilize water magnetization

(blue curve). If water signal is saturated, the performance of the CP scheme deteriorates (red curve). In fact, the efficiency of the CP element becomes comparable to that of the refocused INEPT (black curve). Hence it is mostly the water magnetization which makes the CP element successful, cf. factor (ii) in the above list.

At the same time we note that CP element is less susceptible to the exchange losses than refocused INEPT, corresponding to factor (iii) in the above discussion. This becomes apparent from the comparison of red and black profiles in Figure 3.14. In both simulations there is no influx of magnetization from water, yet CP holds an advantage since it is less affected by the loss of $2H_yN_z$ and $2N_yH_z$. This advantage, however, is not as significant as the gain from water magnetization. Therefore we conclude that it is mainly factor (ii) which makes CP scheme superior, whereas factor (iii) plays a smaller role. This observation holds true for the entire range of k_{ex} rates relevant to our experiment.

3.2.4.2 The Efficiency of CP Scheme: Dependence on k_{ex}

Amide exchange rates in disordered proteins are dependent on the primary sequence. As a result, an IDP sample presents a spectroscopist with relatively broad range of k_{ex} values. If the sample also includes partially or fully folded components, as is the case with drkN SH3, then there is an additional set of signals with very low k_{ex}. Generally one may expect to encounter the situation where the peaks in the spectrum show a range of k_{ex} extending from 0 to $\sim$ 200 s^{-1}. Ideally, all of these peaks should be captured in a single experiment. As it turns out, the CP scheme can effectively deal with this situation: one τ_{mix} setting is sufficient to capture all resonances with near-optimal sensitivity. Specifically, the consensus setting $\tau_{mix} = 10.7$ ms achieves transfer efficiency within 6% of the optimal level over the entire range of k_{ex} rates. The refocused INEPT scheme is somewhat less forgiving, i.e., a single τ_{mix} value cannot accommodate the range of k_{ex} values quite as well.

The range of k_{ex} rates from 0 to $\sim$ 200 s^{-1} is generally suitable for $^1H^N$–^{15}N spectroscopy. In this region the advantage of the CP element over refocused INEPT

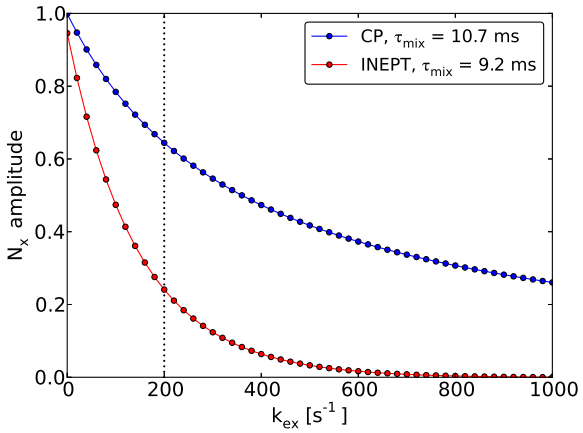


Figure 3.15. The efficiency of CP element versus refocused INEPT transfer as a function of amide $^1H^N$ solvent exchange rate k_{ex}.

is expressed by the average factor 1.5 (see Figure 3.15). The area outside this region is more problematic from the perspective of $^1H^N$–^{15}N spectroscopy. For solvent exchange rates exceeding 500 s^{-1} refocused INEPT becomes completely inefficient, although CP element retains a reasonably good performance. In what follows we investigate the performance of CP element on the sample of Sos at elevated temperature where k_{ex} reaches 620 s^{-1}. It is demonstrated that even under these extreme conditions it is possible to obtain CP transfer with $\sim$ 70% efficiency (which also requires even longer τ_{mix}).

We have also conducted spin dynamics simulations for Lys NH_3 groups. The results were found to be similar to those shown in Figure 3.15, suggesting that CP element should be equally useful in the context of lysine side-chain experiments[140]. As a caveat, note that the maximum transfer efficiency in this four-spin system cannot exceed $\sim$ 44% (this limit holds for both refocused INEPT and CP schemes).

3.2.4.3 The Efficiency of CP Scheme: Dependence on *rf* Field Settings

Two conditions need to be met in order to achieve high efficiency of the CP element: (i) cross-polarization transfer per se should be effective, i.e., amide proton

magnetization should be locked and fulfill Hartmann–Hahn condition, $\nu_1^{\mathrm{H}} = \nu_1^{\mathrm{N}}$, and (ii) water magnetization should be effectively locked along the same axis as amide proton magnetization. Both requirements could be readily satisfied if infinitely strong *rf* fields were available. In practice, however, ν_1^{N} is subject to rather stringent technical limitations, which automatically puts a limit on $\nu_1^{\mathrm{H}}(= \nu_1^{\mathrm{N}})$. As a consequence, ν_1^{H} is actually comparable to the offset Δ between the amide $^1\mathrm{H}^{\mathrm{N}}$ resonances and the water signal.

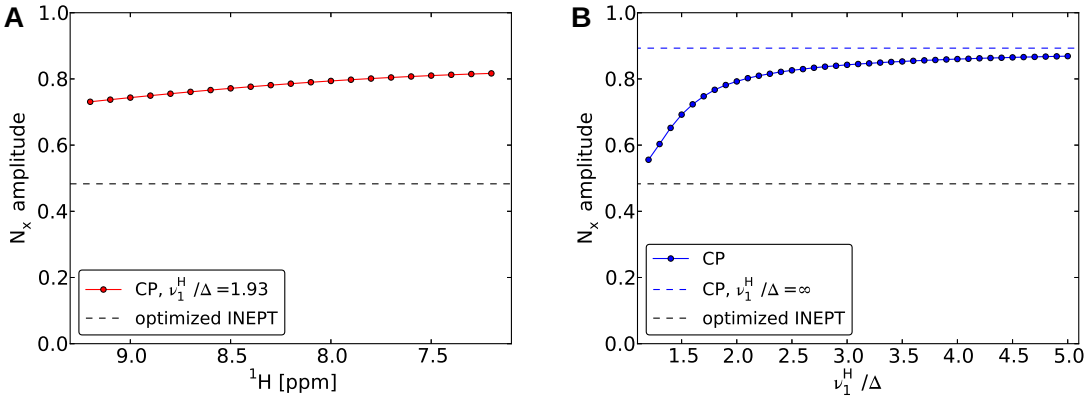


Figure 3.16. The efficiency of CP element for backbone amides as a function of (A) $^1\mathrm{H}^{\mathrm{N}}$ chemical shift and (B) the strength of *rf* field used in CP mixing, $\nu_1^{\mathrm{H}}(= \nu_1^{\mathrm{N}})$, relative to the offset between the proton carrier and the water signal, Δ. The simulations assume DIPSI-2 mixing with $\tau_{\mathrm{mix}} =$ 10.7 ms, proton *rf* carrier positioned at 8.2 ppm, and $\mathrm{k_{ex}} = 100\ \mathrm{s}^{-1}$. The results in (B) additionally assume that proton *rf* carrier is on-resonance with $^1\mathrm{H}^{\mathrm{N}}$ spin.

To address this aspect of the problem we have conducted a series of numerical simulations. In particular, the simulated data shown in Figure 3.16A (red curve) imitate the conditions of our drkN SH3 experiment. In these simulations we assume that proton *rf* carrier is placed in the middle of the amide region, 8.2 ppm, such that the offset relative to that water signal is 3.5 ppm (Δ = 2.1 kHz). We further assume that the strength of the DIPSI-2 *rf* field is ν_1^{H} = 4.0 kHz and the solvent

exchange rate is $k_{ex} = 100\ s^{-1}$. Using this set of parameters we have calculated the CP "excitation profile", i.e., determined the amount of N_x magnetization generated by the CP element for the range of $^1H^N$ chemical shifts. As it turns out, the entire amide region extending from 7.2 to 9.2 ppm is covered in a fairly uniform fashion (red curve in Figure 3.16A). Somewhat higher efficiency is achieved at the upfield end of the spectrum compared to the downfield end, 0.80 vs. 0.72. This observation can be readily understood. Recall that CP transfer is more efficient when amide and water magnetization are both locked along the x axis. This is easier to achieve for $^1H^N$ magnetization precessing at 7.2 ppm than for $^1H^N$ magnetization precessing at 9.2 ppm.

For the problem at hand the following relationship always holds true: $\nu_1^H, \Delta \gg k_{ex}$. Under this condition it can be shown that the efficiency of the CP element depends on the dimensionless ratio ν_1^H/Δ. Figure 3.16B shows how the efficiency of CP transfer changes with the increase of ν_1^H/Δ. It is clearly desirable to use stronger *rf* fields in order to boost the sensitivity of the experiment. The inspection of the simulated data in Figure 3.16B suggests that $\nu_1^H/\Delta \sim 2.0$ should be considered a reasonable target for CP-HISQC measurements. Increasing the strength of *rf* field beyond this level produces only marginal improvement (cf. the curve in the graph which levels off and converges towards the plateau). In our experimental study of drkN SH3 the parameter ν_1^H/Δ was 1.93. The corresponding CP excitation profile is shown in Figure 3.16A. In the study of the Sos peptide ν_1^H/Δ was 2.70 since $^1H^\varepsilon$ resonates closer to the water resonance and hence the offset Δ is smaller.

As for the choice of proton carrier setting. From the perspective of Hartmann–Hahn transfer, it is advisable to set proton *rf* carrier in the middle of the amide region, at around 8.2 ppm. On the other hand, to simultaneously lock amide and water magnetizations one may prefer to place the carrier between the two respective signals, at around 6.5 ppm. The simulation results indicate that 8.2 ppm is very close to an optimal choice. A small improvement can be obtained by slightly shifting the carrier towards water. For $\nu_1^H/\Delta = 2.00$ this improvement becomes inconsequential (0.2%).

Therefore, the standard choice of proton carries setting 8.2 ppm corresponding to the middle of the $^1H^N$ produces near optimal results and can be used in most cases about backbone amide group study.

3.2.4.4 Water Magnetization Preservation Properties in CP-HISQC

For labile protons in our drkN SH3 experiment solvent exchange rates mostly fall in the range 15–150 s^{-1}. On the other hand, the duration of the recycling delay d_1 is 1.0 s. Under these conditions, it is clear that amide $^1H^N$ magnetization is completely replaced with water magnetization during d_1 and therefore the optimal length of the recycling delay is dictated by recovery of water (rather than the relaxation properties of $^1H^N$). Consequently, it is important to prevent water saturation during the pulse sequence, and water magnetization should be well preserved throughout, such as to avoid long recovery periods.

CP-HISQC sequence is designed such as to ensure good water magnetization preservation properties. First, water magnetization is placed along the x axis by hard ^1H pulse. Then it is effectively locked by a train of DIPSI-2 pulses, all of which have phase x or −x. After that water magnetization is returned to the z axis prior to application of the pulsed field gradient. Note that this scheme is subject to certain requirements regarding the synchronization of τ_{mix} with T_{DIPSI2} and the choice of ν_1^H, which have already been discussed earlier. The same tactics is used during the t_1 period, where the water is first placed along x, then effectively locked by the train of (on-resonance) WALTZ pulses with phase x or −x, and after that returned to z axis. Finally, during the refocused INEPT period, the effort is made to keep water magnetization along the z axis. The success of the water-preservation strategy can be conveniently characterized by the so-called water flip-back ratio f, which simply reflects the ratio of water magnetizations after and before each scan[382].

We have determined the value of f for different experimental sequences used in this study according to previous developed procedure[382] (all measurements were conducted on the sample of Sos at 30 °C). Considering 1D version of the experiments

(i.e., t_1 set to zero) we have found that FHSQC and SOFAST-HMQC have flipback ratio f of 97 and 98%, respectively. Given that these sequences nicely preserve water magnetization, they can be executed with very short recycling delays. On the other hand, the HISQC-family sequences do not fare as well, with $f = 0.89$ found in the original HISQC, 0.93 in HISQC-wg3919, and 0.94 in CP-HISQC. These values typically require recycling delays on the order of 1 s. For $f = 0.94$ and $d_1 = 1$ s it is easy to estimate that CP-HISQC experiment reaches the steady state where $\sim 75\%$ of the equilibrium water magnetization are available prior to each scan[382].

The coefficients f also show substantial dependence on t_1. For instance, in the CP-HISQC experiment water magnetization experiences certain loss while being locked by WALTZ pulse train during t_1 period. For long t_1 evolution time, $t_1 = 175$ ms, this loss amounts to $\sim 15\%$. Therefore the f values derived from 1D experiments may not be ideally suited to determine the optimal recycling delays. As an alternative, we have recorded several trial 2D spectra and thus determined the d_1 values which ensure (near) optimal S/N ratio. Based on this standard procedure, the interscan delay was set to 0.2 s for SOFAST-HMQC and FHSQC experiments, while it was set to 1.0 s for HISQC, HISQC-wg3919, and CP-HISQC. All of these settings are consistent with the duty cycle requirements.

3.2.4.5 CP Transfer for Amides with Very Fast Solvent Exchange

Although HSQC-style scheme can be used with the samples where solvent exchange rates k_{ex} approach ~ 200 s^{-1}, however, above this threshold HSQC-type experiments become impractical because of extensive line broadening and loss of magnetization during the coherence transfer steps involved with labile ^1H. This is especially relevant for Lys and Arg side chains, where solvent exchange at pH 7.4, 37 °C is exceedingly fast. Many of the backbone $^1H^N$–^{15}N sites also suffer from very fast solvent exchange under physiological conditions.

In this situation one may opt for an alternative sequence design, different from a regular HSQC-like scheme. Specifically, let us consider a design where magnetization

transfer starts from $^{1}H^{N}$, whereas detection is done on another spin, such as $^{1}H^{\alpha}$ or $^{13}C'$[141,383,384]. In what follows, we demonstrate that CP-based magnetization transfer $H_x^N \rightarrow N_x$ is well suited for use in any of such pulse sequence. In fact, the CP element proves to be indispensable if the spectra are to be recorded in the presence of very fast solvent exchange. Indeed, the simulations shown in Figure 3.15 suggest that CP transfer remains fairly efficient for k_{ex} up to 1000 s^{-1}. This is in stark contrast to INEPT which becomes totally ineffective in this regime.

In order to test these predictions experimentally, we have implemented a specialized pulse sequence for use on Arg side chains. Briefly, the magnetization transfer follows the path $H_x^{\varepsilon} \rightarrow N_x^{\varepsilon}(t_1) \rightarrow C_x^{\delta} \rightarrow H_x^{\delta}(t_2)$. The first transfer step is accomplished by means of the CP element which is identical to the one described in Figure 3.10. The remaining steps are implemented in a standard fashion. The new sequence has been termed (HE)NE(CD)HD (Figure 3.17).

The (HE)NE(CD)HD spectra have been recorded on the sample of Sos peptide at pH 6.0. In addition to the data at 37 °C, we have also collected the data at 44 and 50 °C with the goal to emulate fast solvent exchange conditions observed under physiological pH. Since the signals from four arginine residues are heavily overlapped, a single effective k_{ex} value has been obtained at each temperature: k_{ex} = 170, 330 and 620 s^{-1}, respectively.

Figure 3.18 shows the build-up of (HE)NE(CD)HD signal as a function of CP mixing time. In addition to the experimental data (circles), the plot also contains the simulated profiles (solid lines). These profiles have been calculated on the basis of the independently determined k_{ex} values, as described above. The only adjustment made in Figure 3.18 involves the overall intensity scaling; other than that, no adjustable parameters have been used in generating this plot. We note that the level of agreement between the experimental and simulated data is highly satisfying.

As already pointed out, the shape of the N_x build-up profile in the presence of fast solvent exchange is different from the familiar oscillating curve. The transfer of magnetization from water to amine/amide proton and further on to ^{15}N is controlled

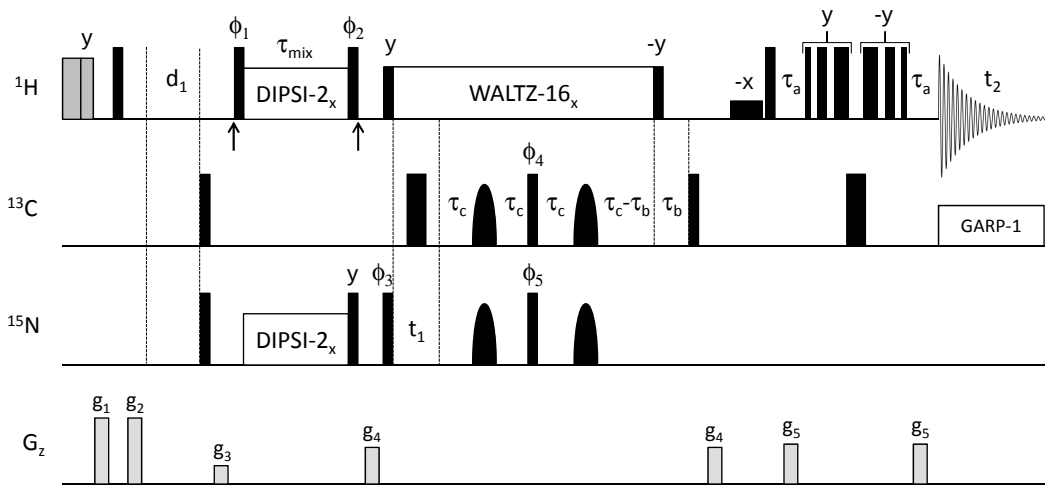


Figure 3.17. (HE)NE(CD)HD pulse sequence for recording ($^{15}N^{\varepsilon}$,$^{1}H^{\delta}$) spectral correlation map for arginine side chains. Proton carrier is set on water resonance except for the duration of the DIPSI-2 period when it is shifted to the middle of the $^{1}H^{\varepsilon}$ region, 7.2 ppm (carrier jumps are indicated by vertical arrows). The delays are $\tau_a = \tau_b = 1.80$ ms, $\tau_c = 22.0$ ms, τ_{mix} variable. The spacing between the consecutive pulses in 3-9-19 WATERGATE[376] is 183 μs. The phase cycle employed is: $\phi_1 = y, -y$; $\phi_2 = -y, y$; $\phi_3 = 2(y), 2(-y)$; $\phi_4 = 4(x), 4(-x)$; $\phi_5 = x$; $\phi_{rec} = x, -x, -x, x, -x, x, x, -x$. Gradient strengths in G/cm (length in ms) are: $g_1 = 20.0$ (3.5), $g_2 = 20.0$ (2.0), $g_3 = 5.0$ (1.0), $g_4 = 10.0$ (1.0), $g_5 = 11.0$ (1.0).

by solvent exchange, which is a stochastic process in essence. As a result, the CP buildup profiles acquire smooth character typical of stochastic processes (e.g., spin relaxation).

Note that in order to obtain efficient CP transfer it is necessary to employ longer τ_{mix}. Considering the data where $k_{ex} = 620$ s^{-1} (red curve in Figure 3.18), the transfer efficiency $\sim 50\%$ is achieved after 15 ms and the transfer efficiency $\sim 70\%$ after 30 ms. This should be borne in mind if CP element is to be used in the presence of ultra-fast solvent exchange, $k_{ex} > 1000$ s^{-1}. In this case the need for long τ_{mix} may constitute a technical limitation because of potential probe heating problem. Note also that in the case of long τ_{mix} the partial loss of water magnetization during this period may not be neglected (due to the relaxation effect of water magnetization itself).

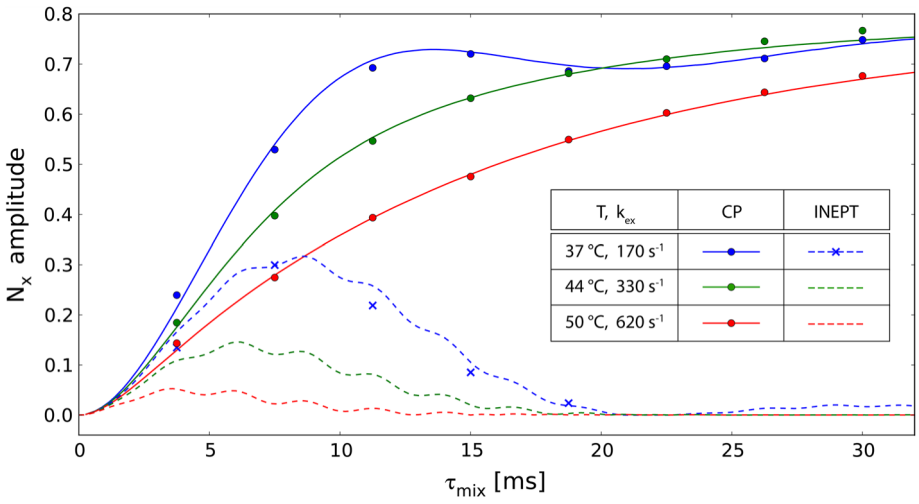


Figure 3.18. The efficiency of CP transfer vs. INEPT transfer in a system with very fast solvent exchange (arginine side-chain $N^{\varepsilon}H^{\varepsilon}$ sites in Sos peptide). The experimental data have been collected using (HE)NE(CD)HD pulse sequence employing proton-to-nitrogen CP element (circles) or, alternatively, refocused INEPT (crosses). The numeric simulation results for CP and refocused INEPT transfer have also been shown (continuous and dashed lines, respectively).

Also shown in Figure 3.18 are the data from (HE)NE(CD)HD experiment where the starting CP element is replaced with a regular refocused INEPT. The experimental data obtained at 37 °C are indicated by blue crosses; the simulated data pertaining to three different temperatures are shown as dashed curves. Clearly, the efficiency of INEPT transfer in the presence of very fast solvent exchange is low. For instance, in the case where $k_{ex} = 620\ s^{-1}$ the efficiency of the INEPT transfer is meager 5% (red dashed curve in Figure 3.18). Obviously, it is not practical to use ^{1}H-to-^{15}N or ^{15}N-to-^{1}H INEPT transfer under very fast solvent exchange conditions.

Measurements targeting arginine, lysine, and histidine side chains, including side-chain sites undergoing very fast solvent exchange, play increasingly prominent role in protein NMR studies[346,385–389]. The use of the CP element in this context is expected to be highly helpful: the fact that CP element still retains its high efficiency in the presence of very fast solvent exchange indicates that this element can be incorporated

into many other pulse sequences to help boost signal-to-noise. For example, many heteronuclei detection experiments developed for studying IDPs totally avoid labile ^{1}H involved in the pulse sequence[141]. However, since ^{1}H has much larger γ than most other heteronuclei (e.g., $\gamma_H \simeq 4\gamma_C$, $\gamma_H \simeq 10\gamma_N$), much better signal-to-noise can be achieved in those experiments if CP element is employed for the initial step of coherence transfer that starts from labile ^{1}H (such as amide group $^1H^N$), and provided that water magnetization is well preserved throughout the pulse sequence.

3.2.5 Conclusion

Hartmann–Hahn cross-polarization in liquids is a highly efficient way of transferring magnetization. It has been shown that CP element has somewhat more favorable relaxation properties than refocused INEPT[381]. CP transfer is also better protected against losses associated with conformational exchange[390]. Furthermore, CP element using DIPSI-2 significantly outperforms refocused INEPT in the presence of substantial *rf* field inhomogeneity[391]. Although it has been recognized in some earlier studies that CP transfer is also more efficient than refocused INEPT for amide sites undergoing fast solvent exchange[392], however, the pulse sequence for general-use HSQC spectroscopy has never been proposed, and in this study we have addressed this omission by integrating the CP element into high-resolution HISQC experiment.

NMR studies of intrinsically disordered proteins are confronted with a number of significant challenges. The main difficulty arises from poor dispersion of chemical shifts in the spectra of IDPs. One should also bear in mind that the samples of IDPs often need to be prepared with low protein concentration because of the IDPs' propensity to aggregate and precipitate. Along the same lines, the measurement time is often limited because the samples tend to quickly deteriorate. If one attempts to conduct the measurements close to physiological conditions, the situation is further exacerbated by fast solvent exchange. Fast solvent exchange leads to intensity loss and dramatic line-broadening in the conventional HSQC-type spectra, making the spectroscopic studies very difficult or impossible.

In this work we have demonstrated that those complications associated with fast solvent exchange can be largely avoided and even turned into an advantage. Proton-decoupled experiment (HISQC) allows one to record the spectra with narrow peak width in ^{15}N dimension. Raising the temperature to (physiologically relevant) 37 °C achieves further sharpening of the signals in ^{15}N dimension, which effectively compensates for exchange-induced broadening in ^{1}H dimension. While the idea of proton-decoupled HSQC experiment (HISQC) is well-known, the benefits of this scheme in the context of backbone $^{1}H^{N}$–^{15}N spectroscopy of IDPs have never received much appreciation.

The fact that water protons rapidly exchange into $^{1}H^{N}$ sites can also be used to one's advantage. First, as has been long realized, fast solvent exchange makes it possible to shorten the recycling delays. Second, as has been found in this study, fast solvent exchange can be used to boost the efficiency of the CP transfer. The CP element is also a natural fit with the proton-decoupled HISQC scheme, thus leading to the CP-HISQC sequence described in this work.

The use of the CP-HISQC scheme does not need to be limited to HSQC-style applications. This sequence can be used as a template for the experiments to measure paramagnetic relaxation enhancements (PREs), residual dipolar couplings (RDCs), etc. In particular, CP-HISQC sequence employs N_x coherence and thus it is perfectly suited for ^{15}N R_2 relaxation measurements. More broadly, the concepts used in CP-HISQC are also applicable to 3D sequences such as HNCO, HNCA, and others. Of interest, it has been found that the properly designed CP element remains efficient even under very fast solvent exchange conditions, $k_{ex} \sim 500$–1000 s^{-1}. This observation suggests that proton-to-nitrogen CP transfer can be used as starting point in the pulse sequences that use non-labile ^{1}H or ^{13}C spins for detection. We have experimentally demonstrated this strategy for $H^{\varepsilon}N^{\varepsilon}$ sites in arginine side chains. Generally we believe that the pulse sequences and concepts presented in this work should signif-

icantly facilitate the spectroscopic characterization of disordered proteins at or near physiological conditions.

3.3 Major Contribution into This Thesis by Other Investigators

Yi Xue provided some helpful suggestions for the NMR pulse sequence design and numerical simulation setup.

CHAPTER 4. PROTEIN–PEPTIDE ENCOUNTER COMPLEX STUDIED WITH CRK–SOS MODEL SYSTEM

4.1 Introduction

The interplay between order and disorder is an essential feature of any proteome. Those proteins in which disorder prevails are known as intrinsically disordered proteins (IDPs). They can be identified with relative ease on the basis of primary sequence: proteins with significant net charge and a low proportion of hydrophobic residues tend to be disordered[393,394]. Additional parameters, such as flexibility and bulkiness of amino acids, can also be incorporated into prediction algorithms[395]. Estimates obtained along these lines suggest that as many as 50% of eukaryotic proteins contain long stretches of disordered residues[396], and more than 20% can be described as predominantly disordered[395]. These statistics demonstrate that IDPs cannot be dismissed as a rare quirk of nature: on the contrary, they represent one of the broad and fundamentally important classes of proteins[395].

Of interest, the proportion of IDPs in eukarya appears to be much higher than in archaea or prokarya (in the case of archaea, only 2% of proteins are predicted to have long disordered regions)[397]. Thus, IDPs can be viewed as an advanced type of molecular machinery that evolved in higher organisms. As it happens, disordered proteins are typically involved in cell signaling and regulation[398,399]. In this context, the unfoldedness of IDPs confers a number of functional advantages. IDPs can literally wrap themselves around their binding targets, thus making use of multiple binding sites[400–402]. This mechanism creates new possibilities in terms of binding affinity and specificity. For example, an intrinsically flexible protein may be able to bind several structurally different partners. Such multitasking helps in the development of efficient, highly integrated signaling networks, in which every protein serves in more

than one role. In addition, IDPs offer the benefit of economy because they present large binding interfaces but do not require large supporting scaffolds[71].

The functional requirements for an IDP binding to its structured target in the context of cell signaling are summarized as below: First, sufficiently high specificity is to be expected of any such interaction. This implies a fairly extensive pattern of contacts between an IDP and its folded target. To form these contacts, IDP needs to adopt an appropriate spatial conformation. Given that intrinsically disordered proteins normally go through a multitude of different conformations, this requirement should effectively slow the binding process and lead to low on rates (k_{on}). At the same time, there is an evolutionary pressure to maintain a high off rate (k_{off}), which is necessary to ensure a rapid response to changing external stimuli. The combination of a low k_{on} and a high k_{off} is expected to produce a large dissociation constant ($K_d = k_{off}/k_{on}$). In a cellular milieu, where the concentrations of proteins are typically low, this should render the entire mechanism unsustainable.

As it appears, nature has found a way to resolve this conundrum. As already pointed out, IDPs usually carry a significant net charge (which often includes a contribution from post-translational modifications, such as phosphorylation)[403–405]. This fundamental characteristic of IDPs makes it possible to rely on the so-called electrostatic steering mechanism to boost k_{on} during the association process[153]. Briefly, long-range electrostatic forces pull an IDP towards the relevant acidic (or basic) patch on the surface of a target protein, thus establishing contact in the vicinity of the binding site. The resulting state, in which IDP is loosely anchored at the periphery of the binding site, has been termed as "electrostatic encounter complex" [146]. This mechanism effectively increases the local concentration of the ligand in the vicinity of the binding site and thus promotes binding. The on rates in excess of 10^7 $M^{-1}s^{-1}$ are invariably associated with this mechanism[155,406].

After the electrostatic encounter complex is formed, it quickly evolves into the final complex, which is structured and usually has a reasonably high affinity. During this transition, the IDP adopts a suitable conformation to make multiple binding

contacts. This process has been described as "folding upon binding" [407], therefore the binding of IDPs to their folded target falls under the rubric of "induced fit" (although the elements of "conformational selection" have also been noted [408,409]). It has also been observed that the individual segments of the IDP "coalesce" around the multiple attachment points on the surface of the target protein [410,411]. The electrostatic interactions continue to play a significant role at this late stage in the binding process. It has been pointed out that electrostatic complementarity is essential for binding specificity in the case of IDP complexes [412]. It has also been suggested that the folding is greatly accelerated by a pattern of charged residues surrounding the binding site [156,413].

In this work we focus on the details of the electrostatic interactions involved in the formation of the electrostatic encounter complex. The combined use of the experimental nuclear magnetic resonance (NMR) data as well as molecular dynamics (MD) simulations allows us to obtain a realistic model for these interactions. Recent advances in computer hardware and MD algorithms made it possible to simulate binding of a small IDP to its folded target using regular (unbiased) MD simulations in explicit solvent. This has been demonstrated for binding of peptides to modular domains such as SH2 [414] and SH3 [198], where the peptides can be viewed as a minimal model for an IDP. At the same time, quite a few NMR studies for low-affinity electrostatic complexes have been carried out mainly based on PRE experiments, such as the study for complex formed by short peptides with the sequence XKKKK and electron transfer protein plastocyanin [150]. In this work, we pursue a synergistic approach to combine experimental NMR data with MD simulations. Of note, we have used a rigorous MD protocol with a state-of-the-art force field, as opposed to coarse-grained simulations [413], replica exchange simulations [415], or Monte Carlo modeling [150,416], which were all successfully applied in the context of peptide docking. The use of *bona fide* MD allows us to draw a direct comparison with the NMR measurables.

For the purpose of this study, we have selected a system consisting of the N-terminal SH3 domain from adaptor protein c-Crk and a 10-residue peptide from Ras

activator protein Sos. Cellular Crk is ubiquitously expressed in a wide range of cells and tissues[417]. In its role as an adaptor protein, c-Crk appears at the center of a vast and complex signaling network: the function of c-Crk is to bring together various proteins that bind to its several modular domains (SH2 and one or two SH3 domains[210,418,419]). In particular, the N-terminal SH3 domain of c-Crk binds a number of guanine nucleotide exchange factors, and these interactions have been directly linked to the high motility of cancerous cells[420]. One of the nucleotide exchange factors that binds to c-Crk N-SH3 is the Ras activator protein Sos[421,422]. The proline-rich sequence that is responsible for this binding belongs to the disordered carboxyl-terminal segment of Sos. The corresponding 10-residue peptide, PPPVPPRRRR, has been crystallized in complex with murine c-Crk N-SH3[197]. The resulting crystallographic structure has also been used in an earlier MD simulation study featuring a shorter version of the Sos peptide, PPPVPPRR[198].

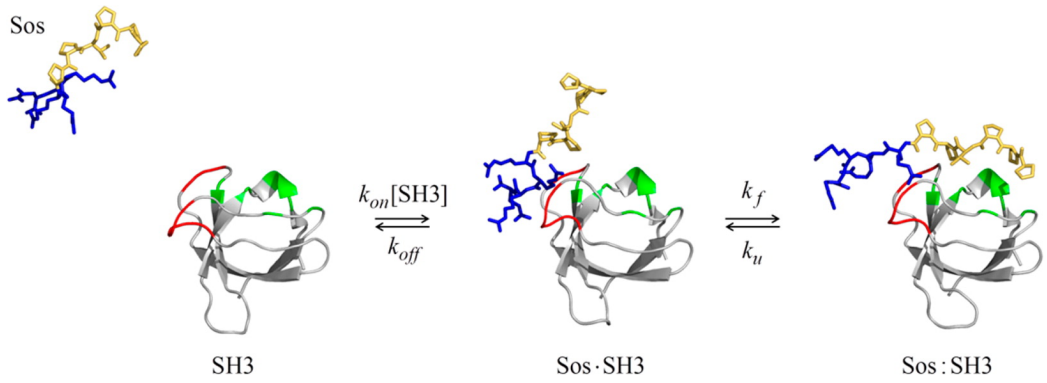


Figure 4.1. Schematic representation of the binding mechanism between the 10-residue peptide Sos and c-Crk N-SH3 domain. The hydrophobic residues in Sos are colored gold, positively charged Arg residues in Sos are colored blue, the negatively charged Glu and Asp residues at the periphery of the peptide binding site on the surface of the SH3 domain are colored red, hydrophobic residues lining up the peptide binding site on the surface of the SH3 domain are colored green.

In the first part of this study, we focus on the interaction between the wild-type murine c-Crk SH3 (wtSH3) and the 10-residue Sos peptide (Sos). The presumed binding mechanism is illustrated in Figure 4.1, which shows the transition from the free state of the peptide (Sos) to the putative electrostatic encounter complex (Sos·wtSH3) and further to the finally formed tight-binding complex (Sos:wtSH3). The formation of the hypothetical encounter complex is underpinned by the interaction between four arginines at the C-terminus of the Sos peptide (blue) and the Glu/Asp-rich patch on the surface of wtSH3 (red). The final complex is then formed as the side chains of the two conserved Sos prolines (gold) are tucked into the shallow hydrophobic grooves on the surface of wtSH3 (green). As a matter of convention, the formation of Sos·wtSH3 is classified as binding (rate constants k_{on} and k_{off}) and the subsequent transition to Sos:wtSH3 is viewed as folding (rate constants k_f and k_u).

The Sos:wtSH3 complex has a relatively high binding affinity ($K_d \sim 1\ \mu M$), therefore it behaves as a tight-binding complex but does not display much encounter complex features. The backbone and side-chain ^{15}N relaxation parameters for the bound form Sos as well as the spectral frequency shifts caused by the binding have been measured. Besides, we have also recorded a long MD trajectory of Sos:wtSH3 in explicit solvent with the Amber ff99SB*-ILDN force field[201–203], and another long MD trajectory representing the free Sos peptide. The MD data have been used to predict NMR parameters, and the calculated results were subsequently compared with the experimental data. This comparison revealed a systematic problem with the MD simulation—as it turned out, the salt bridge interactions in Sos:wtSH3 were consistently overemphasized. This issue has also been recognized in several other recent MD simulation studies based on small molecule model systems[423,424]. To address this problem, we have introduced an ad hoc correction to the MD force field. Specifically, the parameters of the pairwise Lennard–Jones potential between the ionized groups in the Arg and Glu/Asp side chains have been slightly adjusted. Using this modified version of the force field, we were able to successfully reproduce the experimental data for the Sos:wtSH3 complex. Furthermore, we were able to simulate the entire binding

process beginning from free Sos, progressing to the Sos·wtSH3 encounter complex, and ultimately undergoing the transition to form the final binding complex Sos:wtSH3. This MD trajectory also produced consistent results with the experimental NMR data in the part pertaining to the Sos:wtSH3 state.

In the second part of this study, we focus on the electrostatic encounter complex formed between Sos and c-Crk N-SH3. The experimental characterization of encounter complexes involved with IDPs poses a serious challenge. First, they tend to be sparsely populated and short-lived. Second, they are comprised of many interconverting conformational species. Third, they are typically found in rapid exchange with the heavily populated free and/or bound states. NMR spectroscopy, which is the most important experimental technique in this arena, is affected by all of these complications. In principle, encounter complexes involving IDPs can be targeted using those experiments that are highly sensitive to the presence of minor species such as relaxation dispersion[425,426], chemical exchange saturation transfer[427,428] or paramagnetic relaxation enhancements[429]. Yet it still remains extremely difficult to capture the three-state exchange process (such as illustrated in Figure 4.1).

In this situation, we resort to an alternative strategy. Specifically, we have modified the interaction of Sos with c-Crk N-SH3 such as to mimic the electrostatic encounter complex. Towards this goal, we introduced two point mutations in the area of two shallow hydrophobic grooves in the SH3 domain (green ribbon in Figure 4.1). These two mutations, Y186L and W169F, are supposed to abrogate the hydrophobic component of binding that is responsible for formation of the final complex. At the same time, these mutations preserve the electrostatic Arg-to-(Asp/Glu) interactions. The main idea is to shift the thermodynamic balance away from Sos:SH3 tight-binding state and towards the Sos·SH3 electrostatic encounter complex state.

The altered system is designed to boost the population of the electrostatic encounter complex. This opens the door for standard NMR experiments, which do not need to be sensitive to weakly populated species. In particular, in the case of the electrostatic encounter complex between Sos and the double-mutant SH3 domain

(dmSH3), we have recorded HSQC titration data which provide information about the dissociation constant K_d ($\sim$ 200 μM), the backbone and side-chain ^{15}N relaxation rates reporting on the mobility of the Sos peptide, and the corresponding ^{1}H and ^{15}N chemical shift perturbations. Using our adaptation of the Amber force field, we were also able to produce an MD model that proved to be in good agreement with the experimental data from the Sos·dmSH3 system. The combined MD and NMR analyses allow us to learn much details about the structural and dynamic properties of protein–peptide electrostatic encounter complex based on Crk–Sos model system, such strategy can also be utilized for studying many other encounter complexes that are involved with disordered proteins or peptides.

4.2 Overview of Protein–Peptide Complex Formed by c-Crk SH3 and Sos

The crystal structure of the complex between c-Crk N-SH3 and the Sos-derived peptide, PPPVPPRRRR, has already been determined (PDB 1CKB[197]). A schematic diagram of the binding interface is shown in Figure 4.2. The peptide is a class II ligand with consensus sequence PxhPx+ (x is any amino acid, h is a hydrophobic amino acid, and + is arginine or lysine)[430,431]. Over most of its length, the peptide adopts a left-handed polyproline type II helix conformation. The helix has a triangular cross section: residues at the base of the triangle interact with the SH3 domain. The side chains of the two conserved prolines, P2 and P5, are tucked into the shallow hydrophobic grooves on the surface of SH3 (buried surface areas of 487 and 377 Å^2 for the peptide and protein, respectively). The grooves are defined by the side chains of conserved hydrophobic residues in the SH3 domain (represented by rounded boxes in Figure 4.2).

The positively charged C-terminus of the Sos peptide is delocalized and, therefore, partially absent from the crystallographic structure (there is no electron density for R10, R9, or the side chain of R8). The side chain of R7 is observed in the electron density map, even though crystallographic refinement reveals a substantial amount of variability at this site: the R7 guanidinium group forms salt bridges with carboxyl

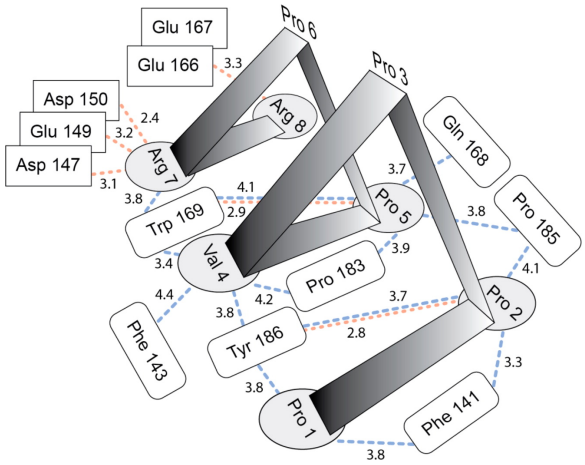
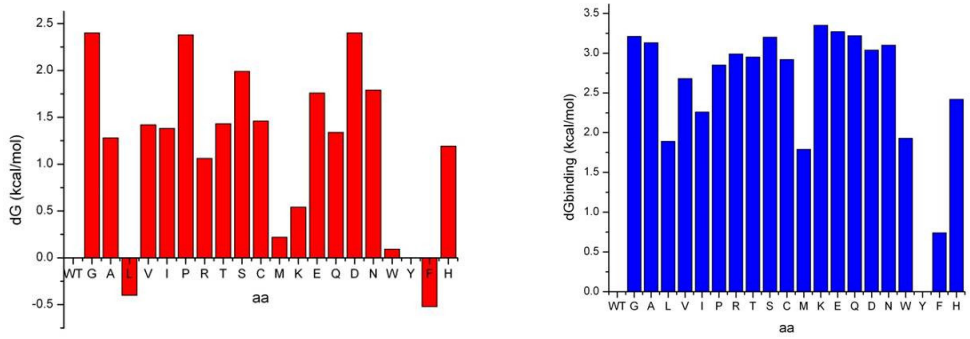


Figure 4.2. Schematic representation of the binding interface formed between Sos peptide and wild-type c-Crk N-SH3 (wtSH3) according to X-ray structure 1CKB[197]. The intermolecular distances indicated in the plot correspond to the nearest pairs of carbon atoms from hydrophobic residues (blue dashed lines) and nitrogen–oxygen pairs in the salt bridges or hydrogen bonds (pink dotted lines).

groups from D150, D147, and E149 (represented by rectangular boxes in Figure 4.2). It can be expected that there could also exist certain amount of interactions between the side chain of R8 with E166 and E167 in a similar fashion[197], although it must be significantly impacted by conformational disorder therefore cannot be directly observed in the X-ray structure. Recall that electrostatic interactions involving the stretch of Arg residues in the Sos peptide anchor the putative encounter complex Sos·wtSH3. Given the dynamic nature of these interactions, this complex cannot be productively studied by X-ray crystallography. This situation calls for the development of new NMR-based methods.

The complex formed between Sos and wild-type c-Crk N-SH3 (wtSH3) displays relatively high binding affinity ($K_d \sim 1\ \mu M$), therefore it is difficult to study electrostatic encounter complex based on this system. However, it has been shown that the population of encounter complex can be much elevated by introducing certain mutations that disrupt the tight-binding property, such modified systems are better

suited for encounter complex studies[432]. The program FoldX[209] is utilized to evaluate the effects of residue mutations on wtSH3. Several hydrophobic residues near the binding interface of wtSH3 (e.g., F141, W169 and Y186) have been selected as mutation targets, since these mutations should not disrupt electrostatic interactions that contribute to encounter complex formation. All possible amino acid substitutions have been scanned at each mutation site, the free energy changes for free SH3 domain (dG) and for the complex formed with Sos ($dG_{binding}$) are evaluated with FoldX. Ideally, the introduced mutation should not affect the stability of SH3 domain itself (i.e., $dG \leq 0$), while it can much perturb the interaction between SH3 and Sos (i.e., $dG_{binding} > 0$). Based on these considerations, several single mutations on wtSH3 have been selected, such as F141W, W169F and Y186L. The FoldX calculation results for Y186 mutation site are shown in Figure 4.3.



(a) dG caused by mutations on Y186

(b) $dG_{binding}$ caused by mutations on Y186

Figure 4.3. FoldX calculation results about the effects of Y186 mutations on wtSH3.

Based on results from FoldX calculations, we have constructed several single-mutant SH3 domains (smSH3) such as Y186L, however, the preliminary ^{15}N relaxation data indicate that the complex formed between Sos and smSH3 still belongs to tight-binding complex but does not display much encounter complex properties. We

continued to construct several double-mutant SH3 domains (dmSH3), finally we chose the one with mutations W169F and Y186L. Our MD and NMR data indicate that the complex formed between dmSH3 and Sos peptide displays many typical features which have also been observed on many other well-studied electrostatic encounter complexes.

4.3 Experimental Results

4.3.1 K_d Determination for Sos:wtSH3 with ITC

ITC experiments were carried out to determine the dissociation constant K_d for the interaction between Sos and wtSH3. Isothermal calorimetry measurements were conducted at 23.4 °C in 20 mM phosphate buffer (same as used for NMR studies) using an iTC200 instrument (MicroCal). The solution of wtSH3 (concentration 0.1 mM, volume 300 μL) was titrated with Sos (concentration 10 mM, injection volume 2 μL, total of 18 injections). The data were fitted using MicroCal Origin, assuming two-state exchange model:

$$\mathrm{Sos} + \mathrm{wtSH3} \underset{k_{OFF}}{\overset{k_{ON}[wtSH3]}{\rightleftharpoons}} \mathrm{Sos{:}wtSH3} \tag{4.1}$$

which is the reduction of the three-state scheme as shown in Figure 4.1. (The notations k_{ON} and k_{OFF} are reserved specifically for the two-state scheme in Equation (4.1), and they are distinct from k_{on} and k_{off} used in the three-state binding scheme) The ITC titration fitting suggests single-site binding with a K_d of 1.5 ± 0.1 μM.

The dissociation constant of $K_d = 5.2 \pm 0.2$ μM for the binding of Sos to wtSH3 in the presence of 150 mM NaCl has been reported before[200]. This result was also reproduced by our fluorescence titration experiments targeting at the intrinsic fluorescence from residue W169 (which resides at the ligand binding interface and thus is sensitive to Sos binding), and a dissociation constant of 5.9 ± 0.2 μM was obtained under identical conditions. On the other hand, our ITC data were collected under

low-salt conditions as used for NMR measurements, therefore, the decrease in the solution ionic strength leads to stronger electrostatic attraction and thus even tighter binding[433]. The affinity of 0.1–10 μM is typical for binding of proline-rich peptides to SH3 domains and, more generally, for binding of peptides to their protein targets[200,434]. Such moderately strong binding is consistent with the signaling function that requires, on the one hand, a meaningful level of complex formation and, on the other, a rapid release of ligand in response to changing stimuli.

The binding affinity is important from the perspective of NMR characterization. Our ^{15}N relaxation measurements and chemical shift determinations have been conducted mostly on samples containing 0.5 mM ^{15}N-labeled Sos and 2.0 mM unlabeled wtSH3. It is easy to verify that under such conditions $> 99.9\%$ of Sos peptide is in bound form. Therefore, the experimental NMR data pertain solely to the bound form Sos peptide.

We also carried out ITC experiments to determine K_d for the complex formed by Sos with dmSH3, however, the obtained ITC titration curve was less than perfect such that it is difficult to extract accurate K_d from the curve fitting. It turns out that this complex has much lower binding affinity than Sos:wtSH3 ($K_d > 100$ μM as determined from NMR titration data). Such low binding affinity locates outside the optimal range for ITC studies, therefore K_d cannot be determined in a reliable fashion[435].

4.3.2 k_{on} and $\Delta\delta_{Sos}$ Determination for Sos:wtSH3 with NMR (Forward) Titration

NMR (foward) titration study was conducted by adding unlabeled wtSH3 stock (3.0 mM) into the sample of ^{15}N-labeled Sos (0.2 mM), and the results are shown in Figure 4.4A. Large titration shifts are observed for residues V4, R7, and R8; for all these sites, the on-off exchange falls in the fast-to-intermediate exchange regime. The line shape analysis of V4 titration yields k_{ON} values of 1.6×10^9 and 2.0×10^9 $M^{-1}s^{-1}$ for ^{15}N and ^{1}H dimensions respectively (see Figure 4.4B). At the same time, the joint treatment of all titration data from V4, R7, and R8 produces the global k_{ON} value

of 1.5×10^9 $M^{-1}s^{-1}$ ($K_d = 1.5$ μM was fixed during the fitting according to results from ITC). Based on K_d and k_{on}, the k_{off} value thus can be determined according to $k_{OFF} = K_d \cdot k_{ON}$, which is $\sim 2 \times 10^3$ s^{-1}. Such approach that ITC data are combined with NMR to obtain a complete kinetic characterization has been commonly used for studying many other protein interactions[436].

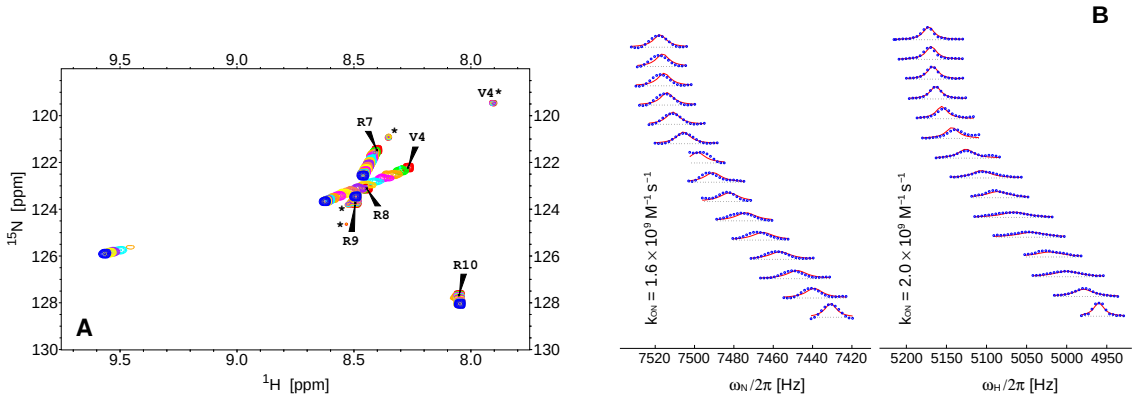


Figure 4.4. ^{1}H–^{15}N HSQC titration of ^{15}N-labeled Sos with unlabeled wtSH3. (A) The series of HSQC spectra recorded during the titration: the starting point corresponds to 0.20 mM Sos without wtSH3 (red contour); the final point corresponds to 0.18 mM Sos and 0.35 mM wtSH3 (blue contour). The minor peaks (labeled with asterisks) represent distinct conformational species arising from *cis–trans* isomerization of proline residues in Sos[437]. (B) Titration of spectral peak V4 (nitrogen and proton dimensions). The traces from HSQC spectra were fitted using LineShapeKin[438].

The extremely high on rate k_{ON} is consistent with the prominent role of electrostatic interactions in bringing together Sos and wtSH3[433,439–441]. In fact, proteins that associate with $k_{ON} > 10^7$ $M^{-1}s^{-1}$ always seem to utilize the electrostatic steering mechanism. Previously, k_{ON} values of up to 5×10^9 $M^{-1}s^{-1}$ have been measured for the binding of barnase mutants to barstar[433]. The k_{ON} value as high as 1.7×10^{10} $M^{-1}s^{-1}$ was found for the complex of the AD2 fragment from the intrinsically dis-

ordered N-terminal transactivation domain of tumor suppressor p53 with the zinc finger 2 domain of the CREB-binding protein[442]. These are extreme cases in which the proteins are propelled towards each other by strong electrostatic interactions.

The k_{ON} value can also be predicted based on atomic coordinates in many different programs. In particular, the program TransComp produces results that are accurate to within one order of magnitude[155]. Although this program has been originally designed for folded proteins, it can also generate reasonable predictions for IDPs[411]. Selected MD frames that represents the approach of Sos to wtSH3 and subsequent binding were used as input for the TransComp calculations. The k_{ON} rate estimated in this manner, $\sim 0.5 \times 10^9$ $M^{-1}s^{-1}$, is in good agreement with the experimental result.

The NMR (forward) tirtation also allows us to quantify ^1H and ^{15}N peak shifts for Sos peptide ($\Delta\delta_{Sos}$) upon binding to wtSH3 (Figure 4.4A). A number of prominent shifts have been observed. In particular, a large down field shift of the amide resonance from R8 (+1.1 ppm on $^1H^N$ and +2.7 ppm on ^{15}N) is due to the hydrogen bond[443] with the side-chain carboxyl group of E166. A sizable up field shift of the signal from the R7 side chain (0.4 ppm on $^1H^{\varepsilon}$ and 1.7 ppm on $^{15}N^{\varepsilon}$) is caused by a ring current in W169, with which R7 forms a cation–π contact. Likewise, the down field shift of the amide signal from V4 (+0.3 ppm on $^1H^N$ and and +1.3 ppm on ^{15}N) is caused by the ring current from Y186. At the same time, residues R9 and R10 show very small binding shifts; apparently, these two residues remain extended into the solvent and are highly disordered. Consistent with crystallographic data, two terminal arginines are only transiently involved in the interactions with wtSH3, although they contribute a lot to the kinetics of Sos–wtSH3 association through the initial electrostatic attraction.

4.3.3 NMR (Reverse) Titration Study for Sos:wtSH3 Complex

The titration experiments between Sos and wtSH3 were also carried out in the reverse fashion, by adding unlabeled Sos stock (10 mM) into ^{15}N-labeled wtSH3 (0.2

mM). A series of ^{1}H–^{15}N HSQC spectra were recorded during the titration process, the results are shown in Figure 4.5.

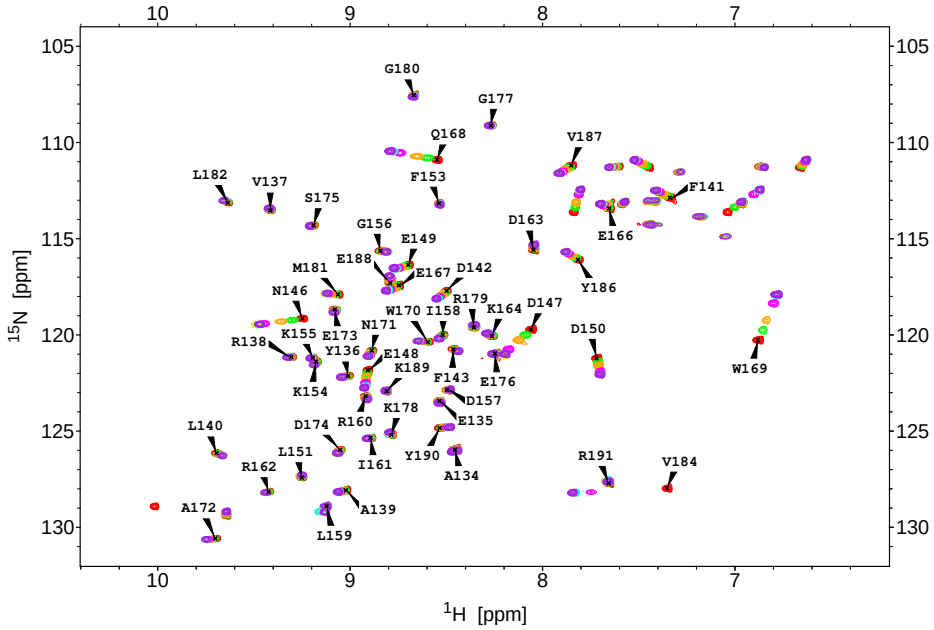


Figure 4.5. ^{1}H–^{15}N HSQC titration of ^{15}N-labeled wtSH3 with unlabeled Sos. The starting point corresponds to 0.20 mM wtSH3 without Sos (red contour); the final point corresponds to 0.18 mM wtSH3 and 1.18 mM Sos (purple contour), such that it finally achieves [Sos]/[wtSH3] $\simeq$ 6.5:1.

The results from the reverse titration experiment basically agree with the information as obtained from the X-ray structure 1CKB—the most significant chemical shift perturbations on wtSH3 ($\Delta\delta_{\mathrm{SH3}}$) came from residues in the vicinity of the binding site (e.g., W169, V184), whereas residues far away from the binding interface display very little peak shifts. The chemical shift perturbation map for wtSH3 generated based on the reverse titration data is shown in Figure 4.6.

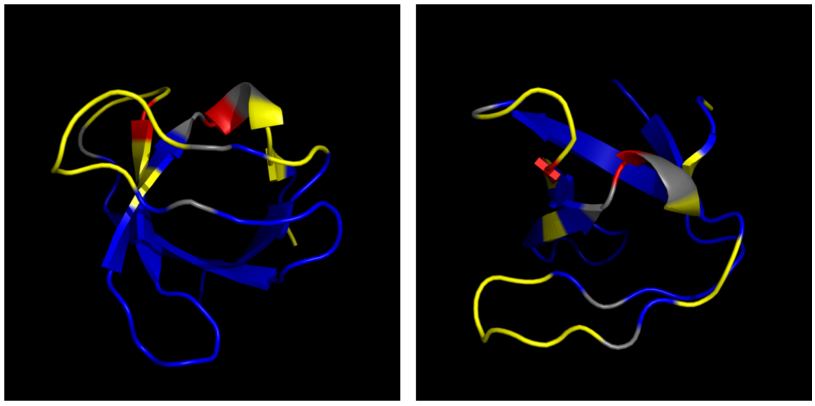


Figure 4.6. Chemical shift perturbation map for wtSH3 binding with Sos. The data are obtained from the pair of ^{1}H–^{15}N HSQC spectra recorded at the beginning or by the end of the reverse titration. The structure is color-coded according to the magnitude of the binding shift $\Delta\delta_{SH3} = \sqrt{\Delta\delta_H^2 + \Delta\delta_N^2}$, where δ_H and δ_N are proton and nitrogen resonance shifts expressed in Hz (for data collected at 600 MHz). The residues with $\Delta\delta_{SH3}$ less than 30 Hz are painted blue, with $\Delta\delta_{SH3}$ between 30 and 150 Hz are painted yellow, with $\Delta\delta_{SH3}$ in excess of 150 Hz are painted red. Four proline residues and residue N144, whose resonance is unobservable and presumably due to extreme exchange broadening, are painted grey.

4.3.4 NMR Titration Study for Sos·dmSH3 Complex

The same series of titration experiments were also carried out for studying the interactions between Sos peptide and dmSH3. In the forward titration experiment, unlabeled dmSH3 stock (2.8 mM) was added into ^{15}N labeled Sos (0.2 mM), the results are shown in Figure 4.7A. Residue R8 still shows the largest amount of binding shift $\Delta\delta_{Sos}$, however, the absolute magnitude of $\Delta\delta_{Sos}$ ($\sim$ 0.1 ppm) is much smaller than the corresponding $\Delta\delta_{Sos}$ as obtained from the titration with wtSH3. Residues V4 and R7 also show certain amount of binding shifts during the titration, but also with much smaller $\Delta\delta_{Sos}$ than the titration with wtSH3. Since in general K_d is correlated with the magnitude of binding shift $\Delta\delta$, these results qualitatively indicate that the binding affinity of Sos·dmSH3 complex is much lower than Sos:wtSH3.

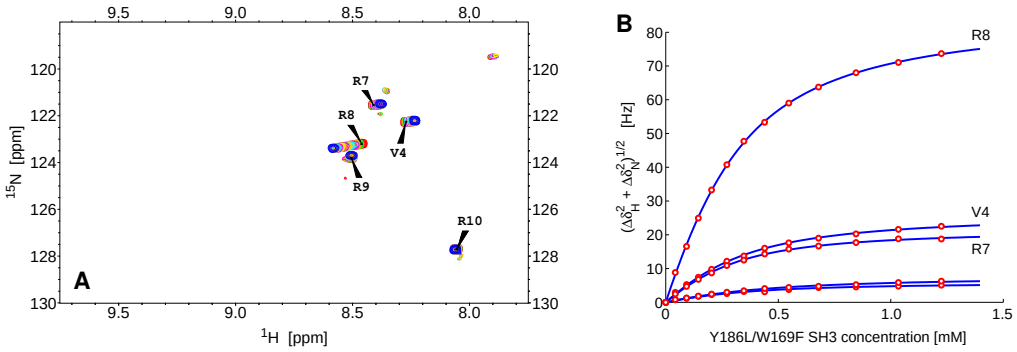


Figure 4.7. ^{1}H–^{15}N HSQC titration of ^{15}N-labeled Sos with unlabeled dmSH3. (A) The series of HSQC spectra recorded during the titration: the starting point corresponds to 0.20 mM Sos without dmSH3 (red contour); the final point corresponds to 0.11 mM Sos and 1.21 mM dmSH3 (blue contour). (B) The fitting of K_d based on binding shifts $\Delta\delta_{Sos}(=\sqrt{\Delta\delta_H^2 + \Delta\delta_N^2})$ from all residues during the titration.

Since all peaks display very little change of the line shape during the titration, it is difficult to extract kinetics data for this complex based on the line shape analysis. However, K_d can be extracted based on the binding shifts $\Delta\delta_{Sos}$ during the titration, which produces $K_d = 192$ μM from the global fitting. Since the strength of electrostatic interactions for forming Sos·dmSH3 or Sos:wtSH3 complexes should be quite similar, it can be assumed that k_{ON} for these two complexes are about the same, such that k_{OFF} for Sos·dmSH3 can be determined as $\sim 3\times10^5$ s^{-1}, which indicates that the average residence time of Sos peptide on the surface of dmSH3 is ~ 3 μs ($=1/k_{OFF}$).

The reverse titration was carried out by adding unlabeled Sos stock (10 mM) into ^{15}N-labeled dmSH3 (0.2 mM), the results are shown in Figure 4.8. The largest binding shifts of dmSH3 ($\Delta\delta_{SH3}$) still come from residues in the vicinity of the binding interface (e.g., V184, F169), however, the absolute magnitude of $\Delta\delta_{SH3}$ on most residues are again much smaller than the titration with wtSH3, therefore it also indicates that the binding affinity of Sos·dmSH3 should be much lower than Sos:wtSH3.

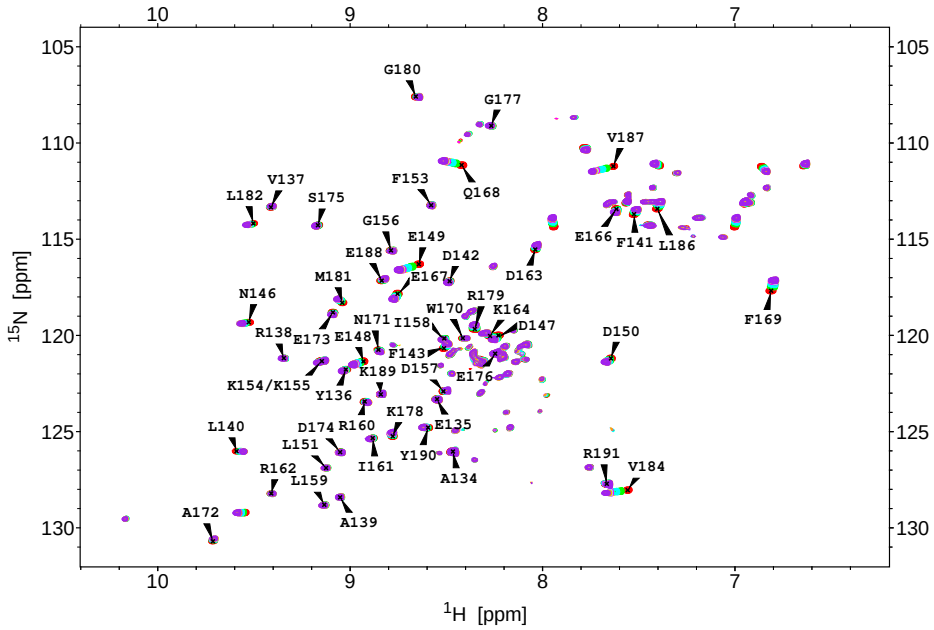


Figure 4.8. ^{1}H–^{15}N HSQC titration of ^{15}N-labeled dmSH3 with unlabeled Sos. The starting point corresponds to 0.20 mM dmSH3 without Sos (red contour); the final point corresponds to 0.17 mM wtSH3 and 1.67 mM Sos (purple contour), such that it finally achieves [Sos]/[dmSH3] $\simeq$ 10:1.

4.3.5 Tumbling Time τ_R Determination for Free wtSH3

The rotational tumbling correlation time τ_R usually needs to be determined experimentally to facilitate relaxation rates calculations from MD trajectories, since it is a well-known issue that τ_R obtained from MD simulations may not be accurate and it is affected by many factor such as the choice of water models[444]. For free wtSH3 in an aqueous solution, the calculation using HYDRONMR[445] predicts a nearly isotropic diffusion tensor with a τ_R of 3.91 ns and a $D_{\parallel}/D_{\perp}$ of 1.05 (see Table 4.1). To experimentally determine the parameters of overall tumbling, we relied on the ^{15}N R_2/R_1 data as analyzed by the program *r2r1_diffusion*[309]. The residues included in the analyses were selected according to previous developed prescription[308] that excludes flexible regions within the protein from τ_R determination. Using the ^{15}N relaxation data from a dilute sample of free wtSH3 (protein concentration of 0.2 mM) and assuming that diffusion tensor is axially symmetric, we obtained τ_R =

4.08 ns and $D_{\parallel}/D_{\perp} = 1.14$, in close agreement with expectations. However, as the concentration of the protein is increased to 2.0 mM, the experimentally determined τ_R value shoots up to 6.64 ns. At the same time, the tumbling anisotropy remains nearly unchanged ($D_{\parallel}/D_{\perp} = 1.18$) and the orientation of the long axis of the diffusion tensor changes only insignificantly, by 16° (Table 4.1). This kind of behavior points towards low-affinity nonspecific (or weakly specific) dimerization, as has been previously observed for a variety of small globular proteins[446–448].

Table 4.1.
Diffusion parameters for wild-type c-Crk N-SH3 in the free form or bound form with Sos peptide[a].

Conditions	τ_R [ns][b]	$D_{\parallel}/D_{\perp}$	θ [°]	φ [°]	N[c]	χ^2_{red}	$\langle R_2/R_1 \rangle$ (protein)	$\langle R_2/R_1 \rangle$ (peptide)
wtSH3 HYDRONMR[d]	3.91	1.05	23	240	53	–	2.62	–
0.2 mM wtSH3 (^{15}N)	4.08/4.08	1.14	14	298	41	0.16	2.73	–
2.0 mM wtSH3 (^{15}N)	6.62/6.64	1.18	26	331	39	0.49	5.31	–
2.0 mM wtSH3 (^{15}N) + 0.5 mM Sos (^{15}N)	6.77/6.77	1.22	29	320	40	0.57	5.51	5.50[e]
4.0 mM wtSH3 + 1.0 mM Sos (^{15}N)	8.69/–							8.34[f]

[a] All ^{15}N relaxation measurements were conducted at 600 MHz spectrometer; the same field strength has been assumed in the HYDRONMR calculations.
[b] Results for isotropic model/axially symmetric model.
[c] Number of residues.
[d] The calculation is based on PDB 1CKB (with Sos peptide deleted).
[e] The data are based on Sos peptide residues V4 and R8, which are part of the well-structured Sos:wtSH3 complex.
[f] The data are based on Sos peptide residues V4, R7, and R8, which are part of the well-structured Sos:wtSH3 complex. These data are also used to calculate τ_R of the complex assuming that the tumbling motion is isotropic.

To quantitatively characterize the self-association effect in wtSH3, we have conducted another $^1H^N$–^{15}N HSQC titration over the range of protein concentrations

from 0.2 to 2.0 mM. A large number of signals are observed to titrate, shifting by a small amount (< 0.06 ppm and < 0.3 ppm in the proton and nitrogen dimensions, respectively). The data have been fitted assuming that the protein undergoes fast exchange between the monomeric and dimeric forms:

$$\delta = p_{\text{mono}}\delta_{\text{mono}} + p_{\text{dim}}\delta_{\text{dim}} \tag{4.2}$$

$$p_{\text{mono}} = K_d(\sqrt{1 + 8(c/K_d)} - 1)/(4c), \quad p_{\text{mono}} + p_{\text{dim}} = 1$$

p_{mono} and p_{dim} are fractions of monomer and dimer respectively, δ_{mono} and δ_{dim} are the corresponding chemical shifts, c is the total protein concentration.

The results of the fitting are presented in Figure 4.9. Briefly, a dimerization site with an affinity constant in the range of 3–8 mM is found in the RT loop. In addition, there are multiple other sites with affinities of > 10 mM that are broadly distributed over the surface of the protein. This is consistent with the notion that at a concentration of 2 mM roughly half of the protein is in the dimeric form, as can be deduced from ^{15}N relaxation data. The dimerization appears to be only weakly specific, which also agrees with the ^{15}N relaxation analyses as mentioned earlier.

4.3.6 Tumbling Time τ_R Determination for Sos:wtSH3 and Sos·dmSH3 Complexes

The τ_R of the wtSH3:Sos complex has been determined using the same sample conditions that were used in the main series of relaxation measurements: 2 mM ^{15}N-labeled wtSH3 with 0.5 mM ^{15}N-labeled Sos. The addition of the peptide resulted only in a small increase in τ_R, ca. 2% (see Table 4.1). Considering that only one-fourth of the protein molecules are loaded with the peptide under the current experimental conditions, we estimate that the τ_R of the Sos:wtSH3 complex is 7.2 ns (assuming an isotropic tumbling model).

Importantly, this result lends itself to direct experimental verification. The sample used to collect R_2/R_1 data was prepared with both wtSH3 and Sos peptide labeled with ^{15}N. Several of the Sos resonances are well-resolved in the spectra, including

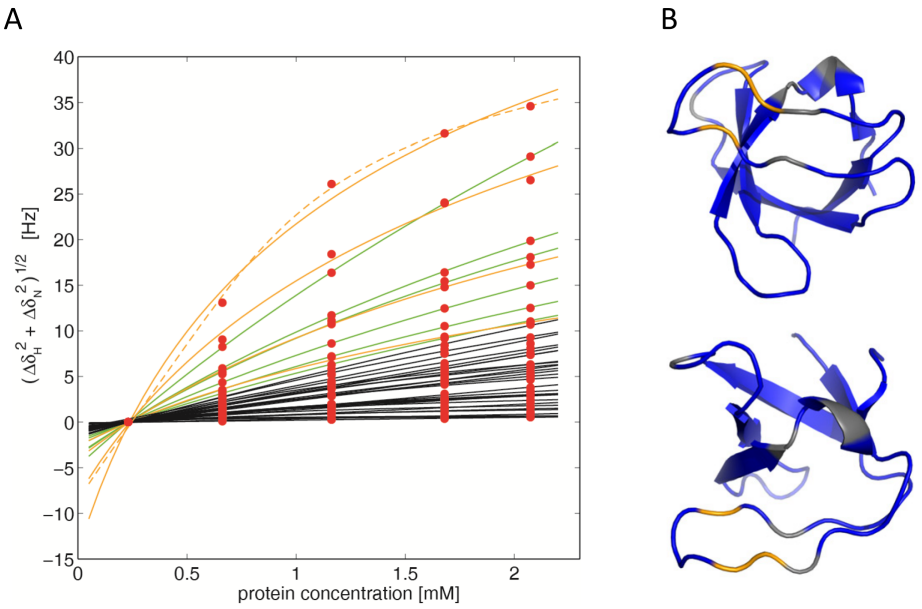


Figure 4.9. Results from ^{1}H–^{15}N HSQC titration that report monomer–dimer equilibrium of wtSH3. (A) Concentration-dependent chemical shifts of wtSH3 during the titration. (B) Structure of wtSH3 highlighting those residues with strongest binding affinities. Painted in gold are four residues with K_d between 3 and 8 mM (corresponding to the gold fitting curves in panel (A)). Painted in grey are four prolines and residue N144, whose resonance is unobservable.

those from V4 and R8 that are in rigid contact with the protein (see the crystallographic structure PDB 1CKB[197]). The R_2/R_1 ratio for these two residues translates into a tumbling time of 6.8 ns for the Sos:wtSH3 complex. More sophisticated model-free analysis using data collected at two magnetic field strengths yields values of 7.1 and 6.8 ns for the two peptide residues at hand. We conclude that the τ_R value of 7.0 ± 0.2 ns represents an accurate and reliable result for the Sos:wtSH3 complex. For the purpose of subsequent MD analyses, it is reasonable to assume that the tumbling is isotropic: the modest anisotropy effect can be safely neglected.

The τ_R for Sos·dmSH3 complex has been determined based on a similar series of samples, mostly focusing on ^{15}N relaxation rates measured for ^{15}N-labeled dmSH3. All results are similar to those as obtained from Sos:wtSH3 complex, therefore we

conclude that this complex also has $\tau_R \sim 7.0$ ns. These results have been used in all subsequent ^{15}N relaxation parameter calculations based on MD trajectories.

4.3.7 ^{15}N Relaxation Parameter Measurement for ^{15}N-labeled Sos Peptide

In order to learn about ps~ns timescale dynamics of the Sos peptide, ^{15}N relaxation parameters have been measured on a series of ^{15}N-labeled Sos peptide samples, in both free form and bound form with wtSH3 or dmSH3.

For the study about Sos:wtSH3, the major focus is on dynamics-modulated electrostatic interactions involving the C-terminal tail of Sos. Accordingly, the NMR data have been measured using the sample of ^{15}N-labeled Sos in the presence of unlabeled wtSH3. The sample was prepared with a 0.5 mM ^{15}N-labeled Sos with 2.0 mM unlabeled wtSH3 to ensure that the spectral signals originate entirely from the bound Sos. Using the standard (^{1}H,^{15}N) experiments, five resolved peaks were observed from the non-proline backbone amide sites in Sos (see Figure 4.4A). In addition, we were also able to record the signals from four (^{1}H$^{\varepsilon}$,^{15}N$^{\varepsilon}$) sites in the arginine side chains, although two of the side-chain resonances are overlapped (R9 and R10).

The standard set of ^{15}N relaxation data, comprised of R_1 and R_2 rates as well as ^{1}H$^{\mathrm{N}}$–^{15}N saturation transfer NOEs (^{15}N–{^{1}H} NOE), has been recorded as described in Materials and Methods. Initially, we reckoned with the possibility that R_2 rates may contain an exchange contribution R_{ex}, although we have not identified any exchange mechanisms that could potentially cause exchange broadening in Sos:wtSH3. From the perspective of MD modeling, the R_{ex} term represents a problem because it normally cannot be simulated (unless an ultralong trajectory is available[183]). To address this issue, we chose to measure transverse cross-correlated relaxation rates η_{xy} that carry the same dynamic information as R_2 but are insensitive to motions on the μs–ms timescale[449,450]. All relaxation data have been collected at two static magnetic field strengths, 500 and 600 MHz. In this work, the comparison between experimental and MD simulation results mainly focus on data measured under 600 MHz. The experimental data from all these sites are plotted in Figures 4.12 and 4.15

as a function of residue number. Panels A–D and E–G show ^{15}N relaxation data for the backbone and side-chain sites, respectively. Panels H–J show the binding shifts $\Delta\delta_{\mathrm{Sos}}$.

The ^{15}N relaxation data for Sos·dmSH3 complex have been measured under the same sample condition (0.5 mM ^{15}N-labeled Sos with 2.0 mM unlabeled dmSH3). However, the low binding affinity of this complex ($K_d \sim 200\ \mu M$) indicates that even four-fold molar excess of dmSH3 still cannot produce $\sim 100\%$ bound form Sos peptide. Based on the determined K_d for this complex, it can be shown that $\sim 90\%$ Sos peptide is in bound form under such sample conditions, therefore ^{15}N relaxation parameters (and also binding shifts $\Delta\delta_{\mathrm{Sos}}$) corresponding to the bound form Sos peptide can be derived based on the following extrapolation strategy (R_1 is shown as an example)[176]:

$$R_1^{\mathrm{exp}} = p_{\mathrm{bound}} R_1^{\mathrm{bound}} + (1 - p_{\mathrm{bound}}) R_1^{\mathrm{free}} \tag{4.3}$$

In this way, we have determined ^{15}N relaxation parameters and binding shifts $\Delta\delta_{\mathrm{Sos}}$ corresponding to bound form Sos peptide within Sos·dmSH3 complex, these results have been used to compare with MD simulation data (Figures 4.21, 4.23). The arginine side-chain region of Sos peptide within Sos·dmSH3 complex produces two groups of peaks with similar intensity in $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ HSQC spectra. Based on the comparison with $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ HSQC spectra obtained from free Sos peptide or bound form within Sos:wtSH3 complex, we conclude that the most reasonable assignment is that one group of peaks come from residues R7 and R8, while another group come from residues R9 and R10.

4.4 MD Simulation Results

4.4.1 Results about Sos:wtSH3 Tight-binding Complex

In this study, the main MD simulation results about Sos:wtSH3 complex are based on the following three MD trajectories:

(i) The 3.06 μs trajectory of Sos:wtSH3 starting from the crystallographic coordinates 1CKB recorded with the standard Amber ff99SB*-ILDN force field, $\lambda = 1.00$. This trajectory is termed MD(wt-xray, $\lambda = 1.00$).

(ii) The 3.20 μs trajectory of Sos:wtSH3 starting from the crystallographic coordinates 1CKB recorded with the altered Amber ff99SB*-ILDN force field, $\lambda = 1.03$. This trajectory is termed MD(wt-xray, $\lambda = 1.03$).

(iii) The 2.28 μs trajectory that represents the binding of Sos to wtSH3 starting from a random configuration. This trajectory has been recorded using the altered Amber ff99SB*-ILDN force field, $\lambda = 1.03$, and is termed MD(wt-rand, $\lambda = 1.03$).

4.4.1.1 Simulation of Sos:wtSH3 Using Amber ff99SB*-ILDN

The trajectory of Sos:wtSH3 has been recorded using the Amber 11 package with the ff99SB force field[201,451]. A number of recent comparative studies, in particular those based on the experimental NMR data, favor this force field over others[193,195,452–454]. In addition, recent corrections for selected backbone[202] and side-chain[203] torsional potentials have been employed. The revised force field is known as ff99SB*-ILDN[455]. The initial coordinates for the MD simulation have been taken from the crystallographic structure 1CKB. The C-terminal residues in Sos that are missing from the crystallographic model have been rebuilt using MODELLER[456]. The system was hydrated using TIP3P water[457]. Given that the C-terminal portion of Sos is dynamic and occasionally becomes lifted off the surface of wtSH3, we opted for a large water box (minimum 14 Å separation between any of the peptide or protein atoms and the water box boundaries). The resulting MD trajectory has a length 3.06 μs. In what follows, this simulation is termed MD(wt-xray, $\lambda = 1.00$). The label indicates that the initial coordinates have been taken from the X-ray structure and the force field parameters have been used as is without any modifications (λ is the scaling coefficient that will be discussed with more details in later sections).

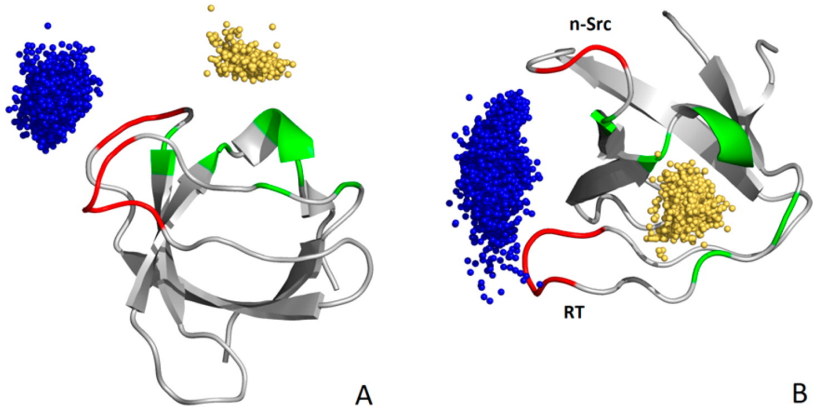


Figure 4.10. Position of the peptide Sos on the surface of wtSH3 according to the data from the 3.06 μs simulation MD(wt-xray, $\lambda = 1.00$). MD frames were aligned by superimposing the SH3 coordinates, the sampling interval for MD frames was 1 ns. The peptide was represented by the center of mass of its N-terminus (residues P1–P6, gold spheres) and C-terminus (residues R7–R10, blue spheres). For wtSH3, the charged residues at the binding interface are colored red (D147–D150, E166 and E167), the hydrophobic residues are colored green (F141, F143, W169, P183, P185 and Y186). Shown are the side view (A) and the top view (B) of the complex.

Figure 4.10 illustrates the dynamics of the Sos peptide bound to the surface of the wtSH3 domain. The gold spheres indicate the instantaneous position of the center of mass of the hydrophobic N-terminal segment of Sos (PPPVPP). The blue spheres correspond to the center of mass of the charged C-terminal segment (RRRR). It is apparent that the N-terminal portion of the peptide is well-localized, whereas the C-terminal portion moves a great deal. The data shown in the graph can be used to calculate the root-mean-square fluctuation for the respective mass centers, which can be conveniently expressed in a form of crystallographic B factors[458]. The values calculated for N- and C-terminal segments are 35 and 209 Å^2, respectively, underscoring the different dynamic status of the two regions.

The top panel in Figure 4.11 shows the root-mean-square deviation (rmsd) trace of the MD trajectory as calculated for the atoms of the Sos peptide using the crystal

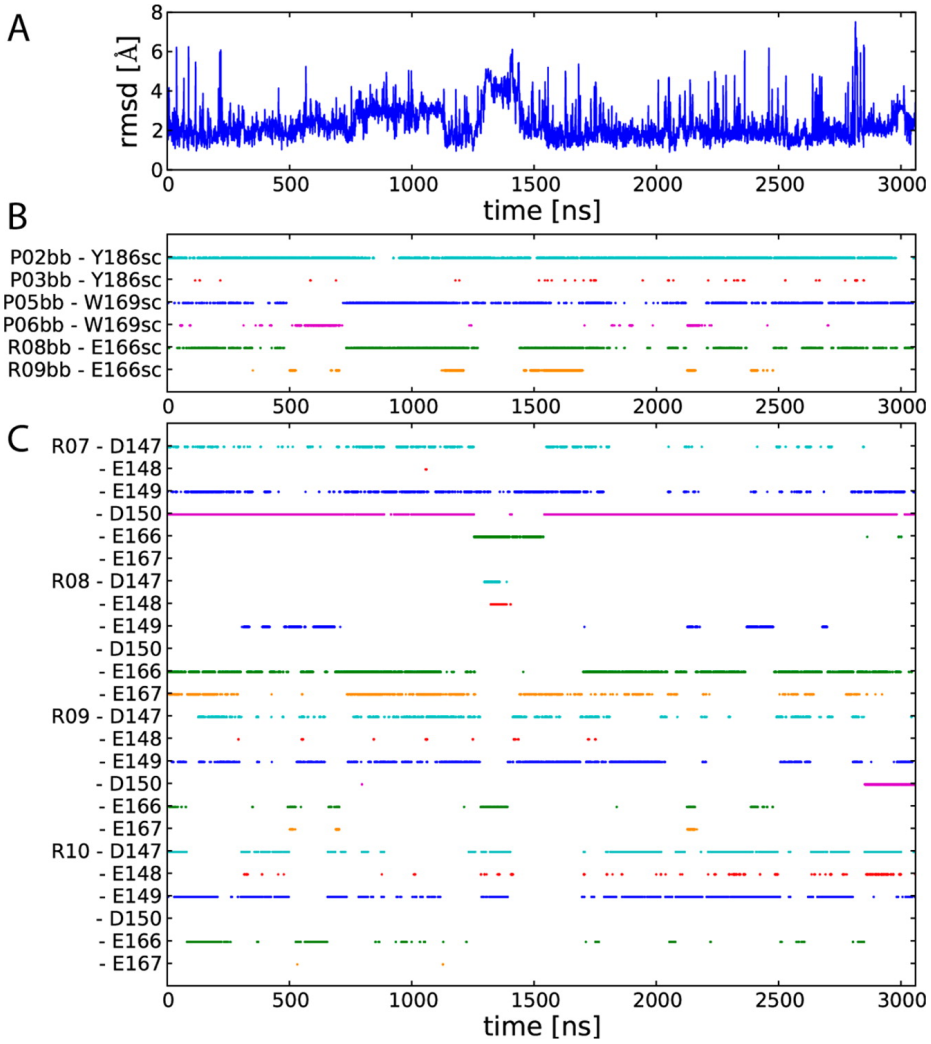


Figure 4.11. rmsd, hydrogen bond and salt bridge properties as calculated from the 3.06 μs MD(wt-xray, $\lambda = 1.00$) trajectory. (A) rmsd of the peptide atomic coordinates relative to the crystallographic structure 1CKB. (B) Peptide–protein hydrogen bonds and (C) salt bridges in the same Sos:wtSH3 trajectory. The trajectory has been divided into 1 ns intervals, and the content of hydrogen bonds or salt bridges has been determined for each interval. If the hydrogen bond or salt bridge is present for more than half of the time during the 1 ns interval, then the dot is placed in the corresponding position in the graph.

structure 1CKB as a reference. These results pertain to the ordered portion of the peptide for which the atomic coordinates are listed in the PDB file. The peptide

clearly maintains the correct binding pose for the entire duration of the trajectory. For almost half of the time the rmsd remains below 2.0 Å, while the average rmsd value is just under 2.3 Å. The pattern of hydrogen bonds and salt bridges as seen in the crystal structure is also preserved in the MD simulation (Figure 4.11B,C). In particular, the signature hydrogen bonds P2–Y186 and P5–W169, as well as the R7–D150 salt bridge, persist throughout the simulation.

At the same time, occasional departures from the crystallographic structure are also observed. For example, at the point in time $\sim$ 500 ns the peptide undergoes a certain amount of twist. The resulting conformation is stabilized by the P6–W169 hydrogen bond that replaces the canonical P5–W169 bond (see Figure 4.11B). This conformation survives for approximately 200 ns, before reverting to the original state. Another notable transition occurs at approximately 1300 ns. At this point in time, residue R8 moves away from E166 and E167 and instead engages D147 and E148. At the same time, the side chain of R7 shifts in the opposite direction, i.e., away from D150 and towards E166. This rearrangement leads to an appreciable increase in the rmsd, reaching $\sim$ 4 Å (see Figure 4.11A).

The dynamic behavior of the charged C-terminal tail of Sos is one of the major interests in this study, since this part is central to the formation of the electrostatic encounter complex. This arginine-rich tail is accommodated in the cleft between the RT loop and the n-Src loop, both of which carry stretches of negatively charged amino acids (see Figure 4.10B). The tail moves mostly in a sideways direction, sweeping over the surface of the protein. In doing so, it sporadically forms hydrogen bonds with wtSH3 as well as multiple salt bridges (see Figure 4.11B,C). Some of these interactions are stable and exist for a significant fraction of time, while others make only a brief appearance. Importantly, at each point in time, the C-terminal tail of Sos is tied to wtSH3 via a significant number of interactions (on average, 4.8 interactions, counting both salt bridges and the relevant hydrogen bonds in panels B and C of Figure 4.11). After a while, the tail rearranges itself, breaking some of its interactions with wtSH3 and establishing a new pattern of contacts.

The experimental data measured for Sos (e.g., $\Delta\delta_{\mathrm{Sos}}$, ^{15}N relaxation rates) offers a good opportunity to put our MD model to a rigorous test. There is a long history of validating MD simulations via spin relaxation rates[453,459,460]. Briefly, the MD data are used to directly compute the dipolar (as well as CSA) temporal correlation functions for the individual amide sites. These correlation functions are subsequently converted into spectral densities, which are in turn used to calculate ^{15}N relaxation parameters. Spin relaxation is directly influenced by fast internal dynamics, i.e., the motions with correlation times on the order of τ_R or shorter, and indirectly by slower motional processes. For example, ^{15}N relaxation can be indirectly affected by rare conformational transitions as seen in Figure 4.11. Indeed, formation and dissolution of salt bridges are expected to have an effect on the fast local dynamics of Sos bound to the surface of wtSH3. From this perspective, it is important that the MD simulation is sufficiently long, 3.06 μs, to sample such relatively rare events.

To calculate binding shifts, we used the knowledge-based prediction program SHIFTX2[461]. The program was first applied to the MD trajectory representing free Sos and then to the trajectory in which Sos is bound to wtSH3, thus allowing for the calculation of $\Delta\delta_{\mathrm{Sos}} = \delta_{\mathrm{Sos}}^{\mathrm{bound}} - \delta_{\mathrm{Sos}}^{\mathrm{free}}$. A number of studies have appeared recently where chemical shift predictors have been applied to the frames from the MD simulations[195,462,463]. Moreover, a new generation of predictors has been developed specifically for this purpose[464,465]. Note, however, that these previous applications have a somewhat different focus—namely, they seek to improve the prediction of chemical shifts by taking into consideration the local dynamics of proteins. In our work, the calculated shifts are used to validate the dynamic scenario seen in the MD simulation.

In particular, $\Delta\delta_{\mathrm{Sos}}$ values are sensitive to the (dynamically modulated) hydrogen bonds, cation–π contacts, and salt bridges such as those formed by several residues in the C-terminal portion of Sos. Hence, the comparison of the calculated and experimental shifts should shed light on these interactions and their dynamic status (degree of motional averaging). Of note, SHIFTX2 employs well-established and reliable algorithms for calculating shifts caused by hydrogen bonds and ring currents.

This makes it well-suited for the purpose of our study, i.e., to probe the interactions between Sos and wtSH3.

In principle, it should also be possible to use $\Delta\delta_{\mathrm{SH3}}$ measured in the reverse titration experiment to further validate the MD data, along the same lines as what has been done with $\Delta\delta_{\mathrm{Sos}}$. In practice, however, we have found that such analysis is of no value. None of the backbone amide groups in wtSH3 forms a hydrogen bond with Sos. Furthermore, Sos contains no aromatic rings. Thus, two of the largest sources of binding shifts are eliminated in this case. As a consequence, the measured $\Delta\delta_{\mathrm{SH3}}$ values are small to moderate: none of the shifts exceeds 0.5 ppm in the $^{1}\mathrm{H}^{\mathrm{N}}$ dimension or 1.3 ppm in the $^{15}\mathrm{N}$ dimension. Both the MD simulation and the SHIFTX2 algorithm are unable to properly capture and process the subtle structural and dynamic changes that cause shifts of this magnitude. As a result, the computed $\Delta\delta_{\mathrm{SH3}}$ values display little or no correlation with the experimental data ($r = 0.40$ for $^{1}\mathrm{H}^{\mathrm{N}}$ shifts, and $r = 0.19$ for $^{15}\mathrm{N}$ shifts). This cannot be construed as evidence against the MD model but rather reflects the limitations of the method.

Figure 4.12 presents the summary of Sos $^{15}\mathrm{N}$ relaxation data and the binding shifts, as measured experimentally (red symbols) and calculated on the basis of the MD trajectory (blue symbols). Let us first focus on the backbone relaxation data from residues V4, R7, and R8. These data are reasonably well reproduced by molecular dynamics. Although certain discrepancies are seen in the R_2 plot (panel B), the agreement is nearly perfect for all other parameters, including transverse cross-correlated relaxation η_{xy} (panel D). As already mentioned, η_{xy} is one of the more accurate measures of protein dynamics that is free of the exchange effect; in terms of information content, η_{xy} is equivalent to R_2. Generally, the backbone amide sites of V4, R7, and R8 can be categorized as tightly bound. They are essentially integrated with the rigid wtSH3 scaffold: their spin relaxation is controlled by the overall tumbling motion of the complex and, to a certain degree, small local fluctuations.

On the other hand, the backbone relaxation data for residues R9 and R10 as well as side-chain data for R7, R8, R9, and R10 show clear systematic deviations

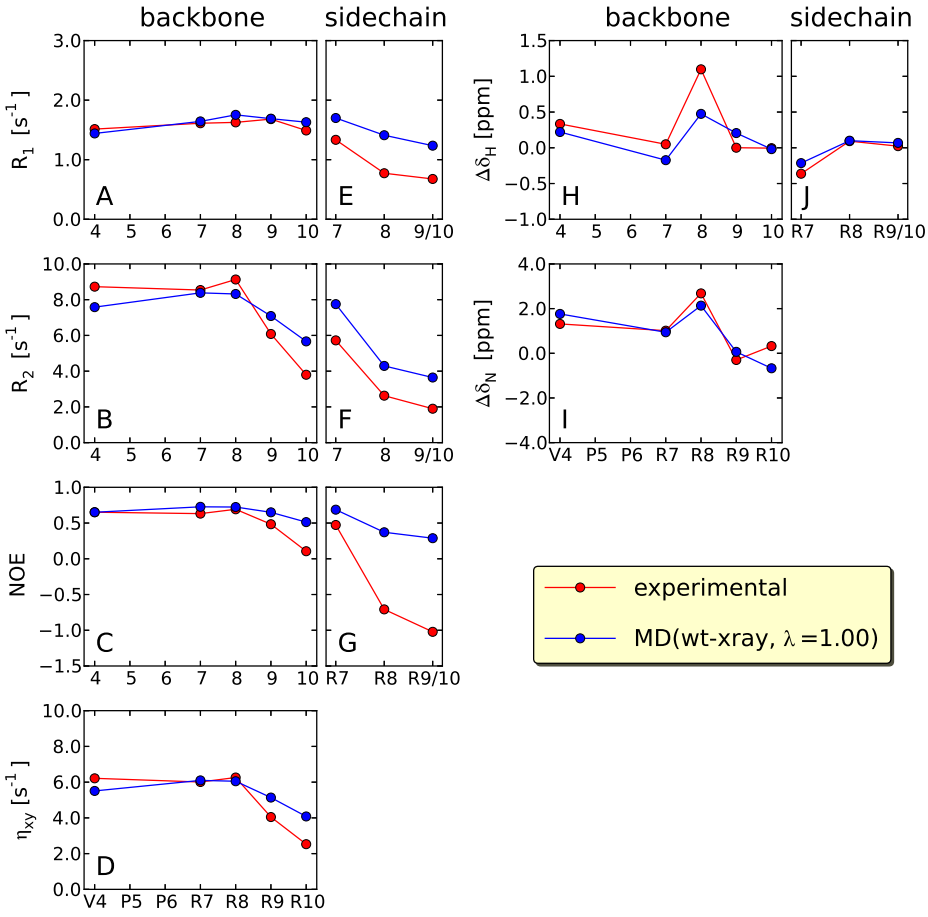


Figure 4.12. Comparison of the experimental (red) and simulated (blue) ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for the Sos peptide in complex with wtSH3 based on MD(wt-xray, $\lambda = 1.00$) trajectory. Simulation MD(wt-xray, $\lambda = 1.00$) uses the original Amber ff99SB*-ILDN force field and starts from the crystallographic coordinates 1CKB. The magnitude of experimental errors is less than or comparable to the size of the symbols.

between the experiment and the MD model (panels A–G). Specifically, the simulated R_2 and η_{xy} rates are higher than the corresponding experimental values. Likewise, the simulated $^{1}H^{N}$–^{15}N saturation transfer NOEs are higher than the corresponding experimental values. This clearly indicates that MD simulation underestimates the amount of local mobility for all of these sites.

According to the MD model, the C-terminal tail of Sos is attached to wtSH3 through the network of salt bridges. While this network is occasionally rearranged, it still has a strong restraining effect on the tail (see Figures 4.10 and 4.11). As a result, the relaxation properties of the tail are rather similar to those of the rest of the complex, which is well-structured and can be viewed as rigid (cf. blue profiles in panels A–G of Figure 4.12). The experimental data, however, paint a rather different picture. It appears that the pair of C-terminal residues in Sos and all the arginine side chains enjoy a degree of motional freedom much greater than that suggested by the MD simulation. This dynamic scenario is also supported by the crystallographic evidence. It is also worth noting that the MD simulation fails to accurately reproduce the large binding shift of $^{1}H^{N}$ in residue R8 (panel H). This observation implicates the side chain of residue E166, which forms an on–off hydrogen bond with this amide as well as transient interactions with several other sites in Sos.

4.4.1.2 Searching for Alternative MD Setup Strategies

In order to find out the root cause of the discrepancy between experimental and MD simulation results on Sos peptide within Sos:wtSH3 complex (especially for C-terminal arginines), we have tested several alternative MD simulation setups. For example, we first tried to use several different water models since the choice of water model is known to affect protein hydration[466–468] as well as conformational preferences and dynamics of disordered peptides[469,470]. Using the Amber ff99SB*-ILDN force field, we have tested TIP3P[457], SPC/E[471], and TIP4P-EW[472] models. In addition, we have also tested a polarizable model SWM4-NDP[473] (supported in the GROMACS package). However, no significant improvement has been observed in any of these trial trajectories, consistent with the recent findings[424]. We have also addressed several other details pertaining to the MD procedure. For example, phosphate ions have been included in the MD setup to reflect the presence of 20 mM phosphate in the NMR buffer. In principle, phosphate ions can form strong interactions with Arg side chains[474], thus competing with the Sos:wtSH3 salt bridges. However, this tactic

also failed to produce any improvements. Likewise, increasing the non-bonded cutoff distance from 8 to 10 Å had no effect on the simulation results. While none of these trials can be considered in any way definitive, we note that our results are consistent with the recent small molecule studies[423,424].

Since all of these alternative MD simulation setups cannot resolve the problem encountered for simulating Sos:wtSH3 complex, we hypothesize that the root cause of the observed discrepancies is due to the deficiency of the MD force field, specifically, the exaggerated strength of the salt bridge interactions[423,424]. In order to demonstrate the validity of this hypothesis, several alternative MD force fields have been tested for simulating Sos:wtSH3 complex.

Two groups of force field parameters are of principal significance for ionic or salt bridge interactions: partial charges and Lennard–Jones parameters. As far as partial charges are concerned, there is a considerable amount of variability between different families of force fields, as well as different generations in the same family. For instance, the charge assigned to N^{η} atoms in the arginine side chain is 0.62 in CHARMM22*, 0.80 in CHARMM27, 0.86 in Amber ff99SB, 0.69 in Amber ff03, and 0.26 in GROMOS 53A6. Besides, there are also significant variations in the Lennard–Jones parameters between different force fields: these variations are largely responsible for dramatic differences in the radial distribution functions associated with salt bridge interactions[423]. In a very recent study which uses small molecules (guanidinium and acetate ions) to model Arg–Asp/Glu salt bridges[424], it has been found that six popular force fields, including CHARMM22*, CHARMM27, Amber ff99SB, Amber ff03, and the newest Amber ff13α force field, all overestimate the population of such salt bridges by a factor of roughly 2. This result essentially does not depend on the choice of water model (TIP3P, SPC/E, TIP4P-EW, etc.).

It can be inferred based on those results as obtained from small molecules that such problems could also exist for MD simulations on proteins, since certain functional groups have very similar properties in small molecules or proteins (e.g., guanidinium–acetate pair also exists within Arg–Asp/Glu residue pair). With this goal in mind,

we have tested AMBER ff03[475] (Amber 11 platform), CHARMM22*[476] (Desmond platform), GROMOS 53A6[477] (GROMACS 4.5.5 platform), and AMOEBA polarizable force field[478] (TINKER 6.0.05 platform using the OpenMM library to support the GPU acceleration). In each case, we have recorded MD trajectories of Sos:wtSH3 using the crystallographic structure as a starting point; the durations of the simulations were 445, 345, 108, and 19 ns, respectively. These trajectories were subsequently analyzed in the same fashion as described above. None of these trials produced the results that we had hoped for. AMBER ff03 shows less propensity to form salt bridges. However, three native hydrogen bonds connecting the backbone of Sos with side chains of wtSH3 are almost completely lost in this simulation, leading to poor predictions of chemical shifts. Similar behavior has been noted for this force field previously[479]. In the CHARMM22* trajectory, two of the native hydrogen bonds are greatly weakened, while arginine side chains remain over-restrained. Chemical shift predictions using this trajectory also prove to be poor. Finally, GROMOS 53A6 significantly underestimates the strength of electrostatic interactions: all salt bridges, including the native R7–D150 bridge, virtually disappear, whereas hydrogen bonds become transient and weak. As a consequence, the bound peptide experiences rapid dynamic fluctuations; in turn, this leads to large systematic deviations between the predicted and experimental ^{15}N relaxation rates.

Of special interest are the results from the 19 ns trajectory of Sos:wtSH3 recorded with the AMOEBA polarizable force field. The polarizability effects are deemed to be important in the context of salt bridge formation[480–482], and polarizable force fields have demonstrated their usefulness in the context of protein–ligand binding and protein structure refinement[483,484]. Although there have been reports of the protein instability in AMOEBA simulations[478], in our case the protein structure remains intact (within 1.5 Å of the crystal coordinates viz. the C^{α} rmsd). However, three native hydrogen bonds linking Sos and wtSH3 are absent in this simulation, and there are only two salt bridges, R7–D150 and R7–D147. Chemical shift predictions have been only partially successful, and there are indications that the C-terminal

segment of Sos remains over-restrained. Given that the simulation is short on the absolute scale, these observations should not be over-interpreted; because of the high computational cost, we have not attempted to extend this AMOEBA trajectory.

Aside from these tests, we also tried to apply the so-called MDEC (molecular dynamics in electronic continuum) model[485–487]. This model focuses on electrostatic interactions involving ionized side chains, which allows modeling the polarization effect in MD simulations in an implicit fashion. In all conventional force fields, ionized side chains are assigned a net charge of $+1$ or -1 and their electrostatic interactions are calculated accordingly. The MDEC model, however, argues that this approach fails to take into consideration the effect of electronic dielectric screening. To remedy this defect, this model suggests to scale the respective point charges with the factor of $\alpha = 1/\sqrt{\varepsilon_{\text{el}}}$, where the electronic (high-frequency) dielectric constant for a protein in aqueous solution is estimated to be $\varepsilon_{\text{el}} \simeq 2$. We have implemented the MDEC strategy in the Amber ff99SB force field and subsequently recorded a number of trial trajectories with different settings of α. The best results were obtained for $\alpha = 0.9$, which is significantly different from the prescribed value ($\alpha \simeq 0.7$). The MDEC approach is admittedly empirical: a rigorous implementation of this concept would require consistent reparametrization of all force field parameters[487]. The rescaling of side-chain point charges should have an effect on many aspects of the system, including side-chain solvation, in a manner that is not easily predictable. Therefore, we abandoned this strategy in favor of a more targeted approach to fine-tune salt bridge interactions for Sos:wtSH3 complex.

4.4.1.3 Fine-tuning Salt Bridge Interactions for Amber ff99SB*-ILDN Force Field

The interaction between two charged atoms is quantum mechanical in nature. When the two atoms are far apart, the interaction energy can be well approximated by the classical Coulomb's law. In contrast, when they approach each other such that the electron clouds start to overlap, the interaction becomes much more complex. In particular, there is an effect of induced polarization that gives rise to attractive

dispersion forces; at short distances, there is also a strong repulsive force due to Pauli interaction. In classical molecular dynamics, all deviations from the standard Coulomb's law are empirically modeled by the Lennard–Jones (LJ) potential (see Equation (4.4)). For the pair of atoms indexed i and j, the potential is defined by two parameters: the interatomic distance at which the LJ energy is at its minimum, r_0^{ij}, and the depth of the corresponding energy well, ε^{ij}. After the atomic charges are set, the LJ parameters determine how close the two atoms can approach each other and are thus strongly related to the magnitude of their optimal interaction energy.

In conventional MD force fields, the pairwise LJ parameters are normally determined by simple combination rules. In particular, each atom type is assigned a van der Waals radius, and the sum of the two radii for any two atoms is taken to be the distance r_0^{ij} for this particular pair. It has been increasingly recognized, however, that in certain cases the simple combination rules require additional fine-tuning to reproduce the desired interatomic interactions. For example, a slight increase in r_0^{ij} between Na^+ (K^+) and Cl^- (over the default value derived from the standard combination rule) made it possible to correct for the otherwise excessive ion pairing in the simulations of simple electrolytes[488]. The LJ parameters for the cation–carbonyl oxygen pairs have also been refined to better reproduce the energetics of K^+ ions in a protein channel[489]. Similarly, a small adjustment in r_0^{ij} has been used to avoid excessively strong binding of the cations to the negatively charged lipid headgroups, leading to an excellent agreement with experimental data[205]. The same approach has been employed to fine-tune the interactions between various metal ions and DNA phosphates, thus avoiding clustering artifacts in the DNA simulations[490]. Other very recent examples include interaction of Mg^{2+} with di- and triphosphate nucleotides[491], as well as interaction between Ca^{2+} and acetate (viewed as an analogue of Asp and Glu side chains)[207]. Finally, pairwise-specific LJ parameters have been implemented for aqueous solutions of alkanes, alcohols, and ethers in the CHARMM Drude polarizable force field, leading to accurate values of the corresponding hydration free energies[204].

Here we use this philosophy to correct for the effect of overstabilized Arg–Asp/Glu salt bridges in Amber ff99SB*-ILDN simulations. Toward this end, a number of trial trajectories have been recorded where the Lennard–Jones r_0 distance for N^{η}–$O^{\delta/\varepsilon}$ pairs has been rescaled. Specifically, we have tested scaling factors λ = 1.01, 1.02, 1.03, 1.04, 1.05, 1.10, 1.15 and 1.20. In each case, the MD data have been processed along the same lines as shown above (Figures 4.10–4.12). Unsurprisingly, large corrections to r_0 led to dramatic attenuation of salt bridges. A good indicator is the native salt bridge between R7 and D150, which is observed in the X-ray structure 1CKB. This salt bridge is present for 86% of the time in the original simulation, λ = 1.00 (trajectory length 3.06 μs), 70% of the time in the altered simulation with λ = 1.03 (3.2 μs), 24% of the time in the altered simulation with λ = 1.05 (0.8 μs), and is completely lost in the simulations with λ = 1.10 or higher. This result, as well as direct comparison of the simulated and experimental ^{15}N relaxation rates and binding shifts, led us to identify λ = 1.03 as the optimal scaling factor. While this is clearly an ad hoc choice, we find that such strategy could also produce better agreement between MD simulation results and experimental data for the encounter complex Sos·dmSH3 formed by double-mutant c-Crk N-SH3 (shown in later sections), which provides as independent validation for such MD force field modification strategy.

It is worth noting that LJ distance parameters have been reoptimized together with partial charges in the new Amber force field ff13α. In particular, the r_0 distance for N^{η}–$O^{\delta/\varepsilon}$ pairs increased by 6%[492]. This new force field remains a work in progress: in its current state, Amber ff13α still shows a tendency to significantly overestimate the strength of salt bridge interactions[424]. Besides, another latest version of Amber force field ff14ipq also suggests that LJ parameters should be fine-tuned for each specific MD simulation as a general routine[493], therefore it is clear that any solution to this problem is likely to involve the Lennard–Jones parameters.

Finally, it is worth mentioning that MD simulations of proteins using implicit solvent are also faced with the problem of overstabilized salt bridges. This problem has

been solved by adjusting the Born radii of the atoms that form the salt bridge[494–496]. Conceptually, this approach is similar to that described above.

4.4.1.4 Simulation of Sos:wtSH3 Using Modified Version of Amber ff99SB*-ILDN

The calibration of Lennard–Jones parameters for N^{η}–$O^{\delta/\varepsilon}$ pairs, as discussed in the previous section, favors the distance r_0^{ij} that is 3% longer compared to the original setting. The new r_0^{ij} value has been coded into the ff99SB*-ILDN force field, thus overriding the standard combination rule for these particular off-diagonal LJ terms. Using the amended force field, we have recorded a 3.20 μs trajectory of Sos:wtSH3 beginning from the crystal coordinates. In what follows, this simulation is referred to as MD(wt-xray, $\lambda = 1.03$) according to the choice of scaling factor λ. Aside from the change in the force field, all other details of the MD protocol were exactly the same as in the original 3.06 μs trajectory, MD(wt-xray, $\lambda = 1.00$). The new simulation is analyzed using the same format as before; the results are shown in Figures 4.13–4.15.

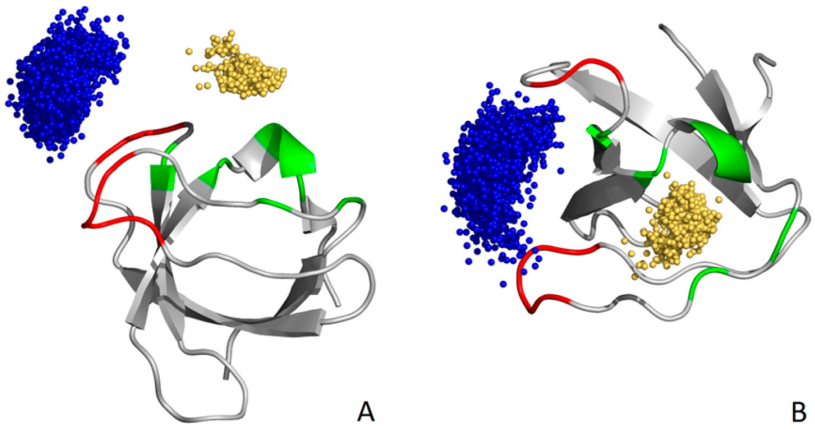


Figure 4.13. Position of the peptide Sos on the surface of wtSH3 according to the data from the 3.20 μs simulation using the modified force field, MD(wt-xray, $\lambda = 1.03$). The plotting conventions are the same as in Figure 4.10.

Figure 4.13 illustrates the fluctuations of the Sos peptide when attached to the canonical binding site on the surface of wtSH3. Gold and blue spheres in the graph represent the centers of mass of the hydrophobic N-terminal segment of Sos and the arginine-rich C-terminal segment, respectively. The results are rather similar to those previously obtained from the conventional simulation (cf. Figure 4.10), although one can also notice certain differences. In the new trajectory, the hydrophobic portion of Sos is somewhat better localized in the binding site. The scatter of the gold spheres in Figure 4.13 corresponds to a temperature factor of 31 $Å^2$, which is slightly lower than the previously obtained value of 35 $Å^2$. At the same time, the arginine-rich portion of the peptide acquires a somewhat greater degree of motional freedom, 229 $Å^2$ vs 209 $Å^2$.

These observations lead us to conclude that there is a certain amount of strain in the Sos:wtSH3 complex. Specifically, formation of strong salt bridges by the C-terminal portion of Sos has a certain destabilizing effect on the binding of the N-terminal portion of Sos. Conversely, weakening of the salt bridges in the $\lambda = 1.03$ simulation creates conditions for tighter binding of the N-terminal segment. This kind of behavior is also manifested in Figures 4.11A and 4.14A. Shown in these graphs is the rmsd trace of the Sos peptide, calculated for the structured portion of Sos (those atoms for which the crystallographic coordinates are available). The graph from MD(wt-xray, $\lambda = 1.03$) simulation clearly displays less dynamic fluctuations: in particular, there are no large-amplitude long-lived fluctuations such as those observed in the MD(wt-xray, $\lambda = 1.00$) trajectory. The average Sos rmsd in the new simulation is 2.1 Å, somewhat lower than that registered previously, 2.3 Å. Hence, reducing the strength of salt bridges appears to enhance peptide binding, which is associated primarily with the well-structured N-terminal portion of the peptide.

An interesting picture emerges from these observations. The electrostatic attraction associated with Arg–Glu/Asp pairing pulls the peptide towards its target, thus efficiently increasing the k_{on} rate. However, once the peptide is bound the same electrostatic interactions involving the C-terminal portion of Sos may no longer be

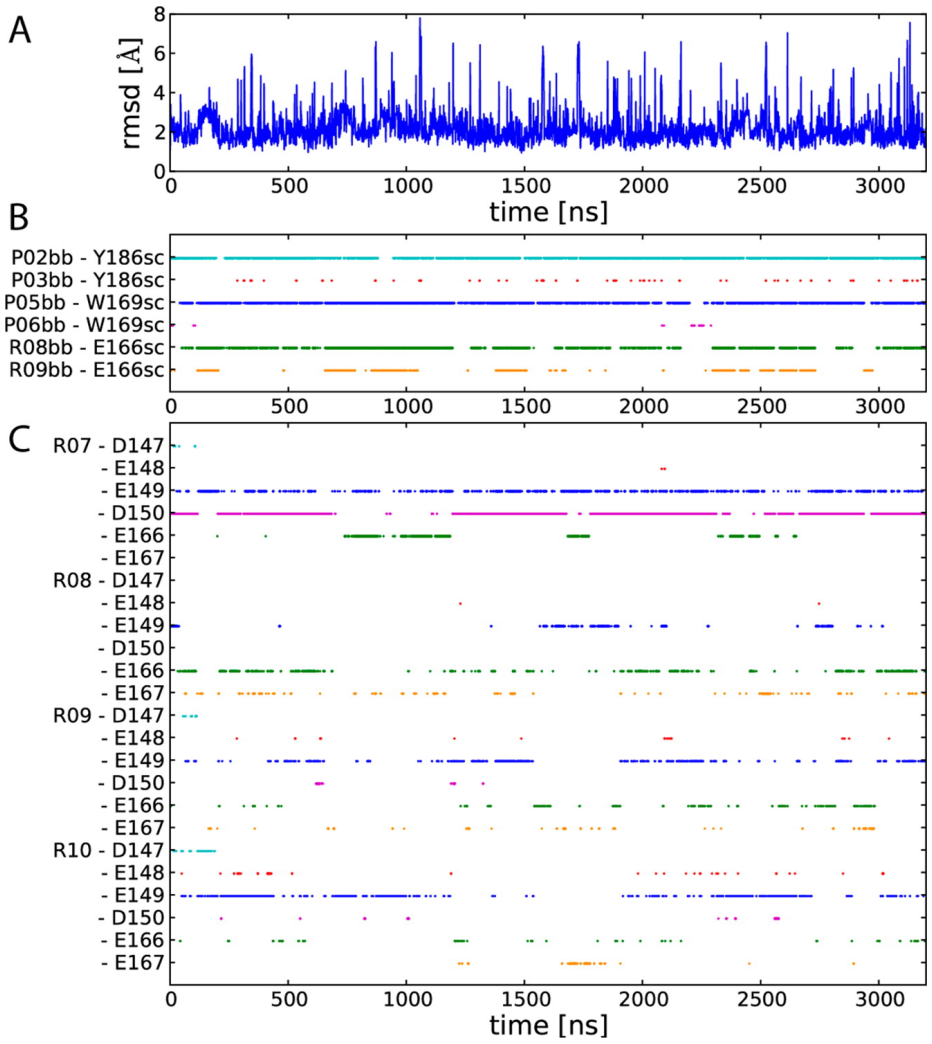


Figure 4.14. rmsd, hydrogen bond and salt bridge properties as calculated from the 3.20 μs MD(wt-xray, $\lambda = 1.03$) trajectory. (A) rmsd of the peptide atomic coordinates relative to the crystallographic structure 1CKB. (B) Peptide–protein hydrogen bonds and (C) salt bridges in the same Sos:wtSH3 trajectory. All conventions are the same as in Figure 4.11.

conducive to binding because these interactions are not necessarily compatible with hydrophobic contacts and hydrogen bonds formed by the N-terminal portion of the peptide. In other words, the propensity to form salt bridges may cause an increase in k_u (see Figure 4.1), thus slightly degrading the binding affinity. This highlights

the dual role of electrostatic interactions in the context of peptide binding: they are responsible for dramatic gains in the on rate, but they may also have a certain destabilizing effect on the bound state.

The traces shown in panels B and C of Figure 4.14 illustrate the time dependence of hydrogen bonds and salt bridges connecting Sos and wtSH3 throughout the course of the MD(wt-xray, $\lambda = 1.03$) simulation. Obviously, the propensity of the C-terminal tail to form salt bridges is greatly diminished. On average, at any given point in time, the C-terminus forms 2.3 salt bridges. This is a 2-fold decrease compared to the original simulation, where the corresponding number was 4.3. At the same time, the hydrogen bond contacts are somewhat strengthened, 2.1 in the altered simulation versus 1.7 in the original simulation. As already discussed above, the reduced propensity to form salt bridges removes some of the steric strain, which, in turn, leads to better hydrogen bonding.

Another way of looking at it would be to determine the number of contacts formed by the structured portion of the peptide (for which the crystallographic coordinates are available) and the remaining unstructured segment. For the structured portion of Sos, the average number of contacts is 3.0 as determined from the MD(wt-xray, $\lambda = 1.03$) trajectory. This includes the native salt bridge R7–D150, which is largely preserved in the new simulation, see Figure 4.14C. The result is quite similar to the one obtained from the original trajectory MD(wt-xray, $\lambda = 1.00$), where the average number of contacts is found to be 3.1. At the same time, the data from the unstructured portion of Sos display some significant differences. The altered MD simulation predicts, on average, 1.3 contacts in this area (salt bridges as well as hydrogen bonds), whereas the original trajectory predicts 2.9 such contacts.

Similar conclusions can be drawn from the analysis of the solvent accessible surface areas. Considering the N-terminal segment of Sos, the average buried surface area is essentially the same in the two simulations, 546 and 544 Å^2. At the same time, there is a significant difference in how the C-terminal segment is packed against the protein

surface—the packing is less tight in the modified simulation protocol, 439 Å^2 vs 574 Å^2.

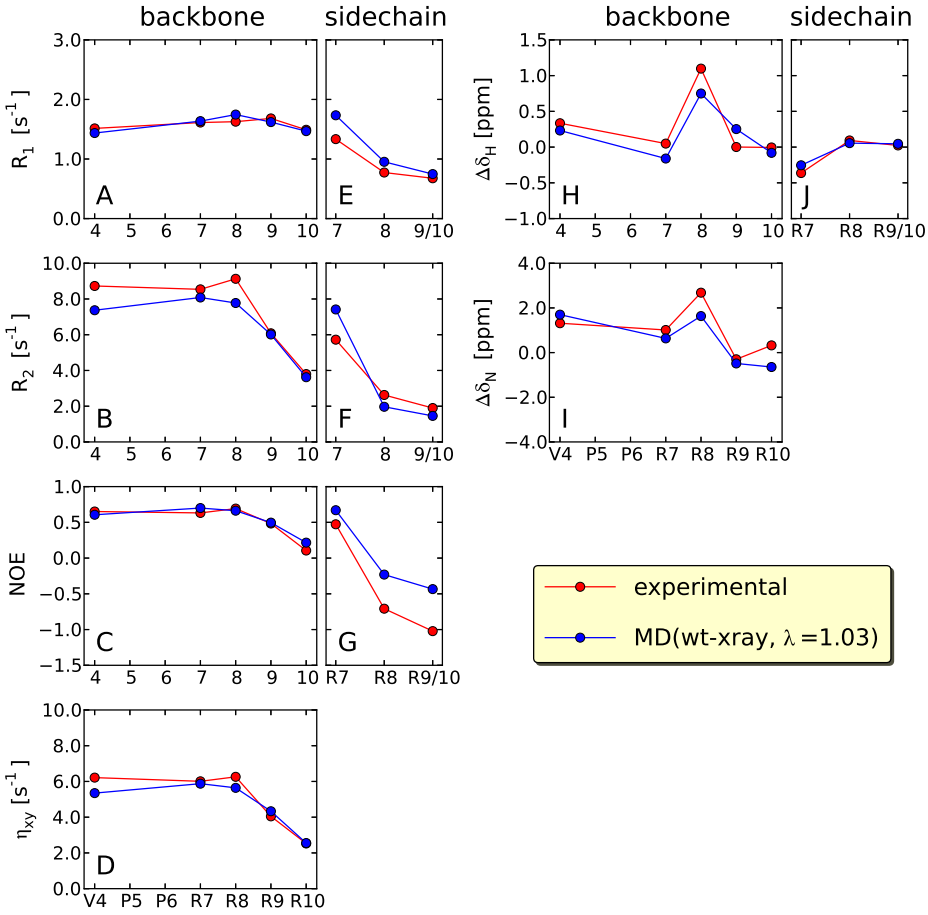


Figure 4.15. Comparison of the experimental (red) and simulated (blue) ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for the Sos peptide in complex with wtSH3 based on MD(wt-xray, $\lambda = 1.03$) trajectory. The MD(wt-xray, $\lambda = 1.03$) simulation uses the modified version of the Amber ff99SB*-ILDN force field in which the strength of salt bridge interactions involving Arg and Glu/Asp side chains has been adjusted.

Finally, the most important element of the validation scheme is the direct comparison between the MD-derived NMR parameters and the experimentally measured data (Figure 4.15). A brief survey of these results indicates a big improvement compared

to the standard implementation of molecular dynamics (Figure 4.12). In particular, the relaxation rates of the backbone amide groups belonging to residues R9 and R10 are now in perfect agreement with the experiment. The arginine side-chain results for residues R8 and R9/R10 are also greatly improved. Furthermore, there is also a noticeable improvement in binding shifts: specifically, the large shift experienced by $^{1}H^{N}$ from residue R8 is now successfully reproduced by the MD-based calculations. At the same time, for those sites that belong to the well-structured portion of the peptide, e.g., V4, the level of agreement between the simulations and experiment remains unchanged (cf. Figures 4.12 and 4.15).

The structure and dynamics basis for the improvement observed in Figure 4.15 can be well understood through the dipolar correlation functions $g(\tau)$ underlying the ^{15}N relaxation rates. The full complement of these correlation functions, for both backbone and Arg side-chain sites in Sos, is shown in Figure 4.16. The plots in the first and third columns show the correlation functions $g(\tau)$ computed to 80% of the trajectory length. In each case, $g(\tau)$ are supposed to converge to a plateau, reflecting the long-term averaging properties of the particular ^{1}H–^{15}N vector. Although the convergence behavior is far from perfect, the data are generally sufficient to judge the amount of disorder at each site. As far as we can tell, the plateau values derived from MD(wt-xray, $\lambda = 1.00$) and MD(wt-xray, $\lambda = 1.03$) trajectories are similar within the uncertainty associated with incomplete convergence (cf. blue and red curves, first and third columns in Figure 4.16). Consider, for instance, the data for residues R9 and R10. The backbone and side-chain correlation functions in R9 and R10 essentially decay to zero. This means that in both trajectories the respective proton–nitrogen vectors experience complete orientational averaging.

The observations concerning the plateaus of the correlation functions (i.e., the order parameters) lead us to conclude that the amplitudes of internal motions in MD(wt-xray, $\lambda = 1.00$) and MD(wt-xray, $\lambda = 1.03$) trajectories are actually similar. This conclusion agrees with the analysis illustrated in Figures 4.10 and 4.13, where the differences between the two trajectories are relatively minor. Therefore, it is not

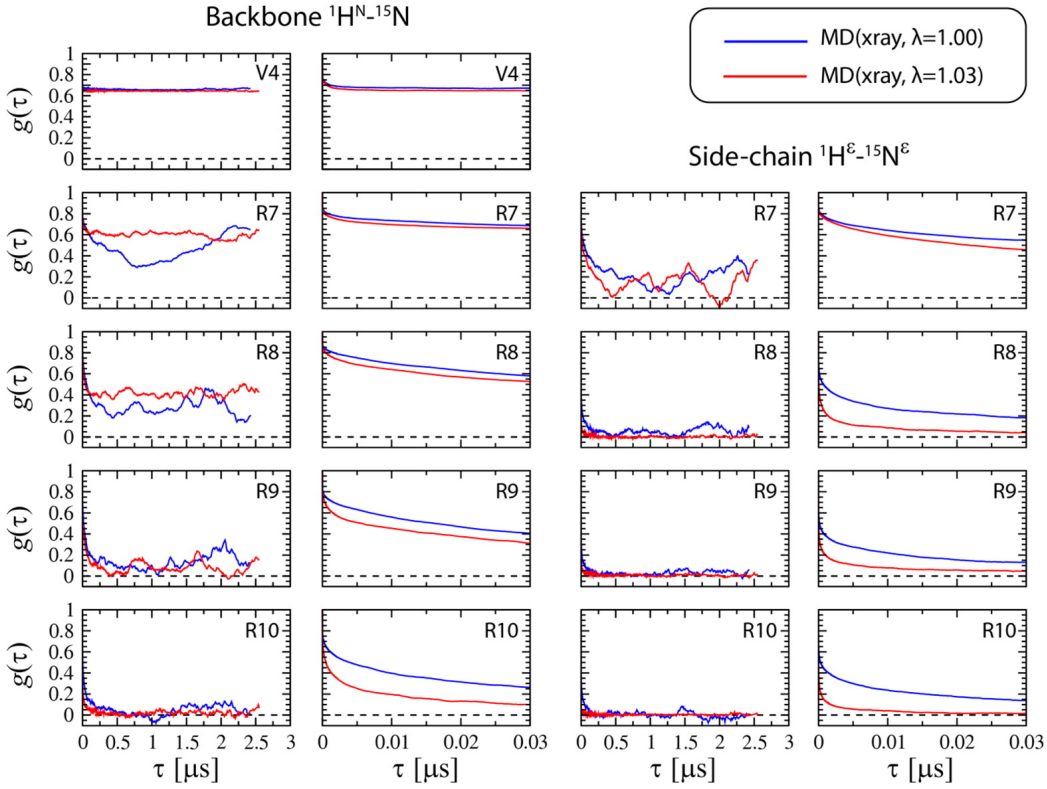


Figure 4.16. ^{1}H–^{15}N dipolar correlation functions $g(\tau)$ for backbone and arginine side-chain sites in Sos as extracted from MD(wt-xray, $\lambda = 1.00$) and MD(wt-xray, $\lambda = 1.03$) trajectories of Sos:wtSH3 (blue and red curves, respectively). The correlation functions have been calculated to 80% of the trajectory length. The expanded view of the initial portion of $g(\tau)$, which is relevant for spin relaxation, is shown in the second and fourth columns.

the motional amplitudes that cause the differences in the simulated ^{15}N relaxation rates. Rather, it is the time scale of internal dynamics that plays the key role. The panels in the second and fourth columns of Figure 4.16 illustrate the expansion of the simulated correlation functions to 30 ns. This is the time interval that is relevant for spin relaxation in the Sos:wtSH3 complex where the tumbling time is $\tau_R = 7$ ns. From these expansion plots, it is obvious that the internal dynamics in Sos is faster in the altered MD(wt-xray, $\lambda = 1.03$) simulation than it is in the original MD(wt-xray, $\lambda = 1.00$) simulation. This observation is relevant for C-terminal residues, R8–R10;

the effect is particularly striking for arginine side-chain sites (fourth column in Figure 4.16).

The analysis described above suggests the following picture of the simulated Sos dynamics. In the MD(wt-xray, $\lambda = 1.00$) trajectory, the C-terminal segment of Sos samples different conformations, where it often becomes trapped for extended periods of time because of the overstabilized salt bridges. In contrast, in the MD(wt-xray, $\lambda = 1.03$) simulation, Sos tends to interconvert between different conformations more rapidly. Consequently, this model is more successful in reproducing the experimental ^{15}N relaxation rates.

Finally, valuable insight into internal dynamics can be obtained from relaxation data collected at multiple magnetic field strengths. With this goal in mind, we have measured the ^{15}N relaxation in Sos:wtSH3 also at 500 MHz field strength. The experimental results are summarized in the left column of Figure 4.17, where solid and dashed lines represent 600 and 500 MHz data, respectively. The right column shows the equivalent data calculated on the basis of the MD(wt-xray, $\lambda = 1.03$) trajectory. Insofar as the field dependence is concerned, the agreement between the experiment and the simulation is very convincing. For relatively rigid sites such as V4, the outcome is dictated mainly by the overall tumbling time, $\tau_R = 7$ ns. However, in the case of flexible residues such as R10, the results reflect a shorter correlation time, corresponding to the internal peptide dynamics (cf., for example, panels D and H in Figure 4.17).

4.4.1.5 Simulations of the Binding Process Between Sos and wtSH3

It has been shown in some earlier study that it is possible to use MD simulations in explicit solvent to reproduce peptide binding in a system that is very similar to Sos:wtSH3[198]. Here we set out to reinvestigate this problem using the modified ff99SB*-ILDN force field ($\lambda = 1.03$) and significantly longer simulations. For this purpose, we have implemented a special protocol to select the initial conformation of the peptide randomly from a 2.1 μs trajectory of free Sos. The peptide was then placed

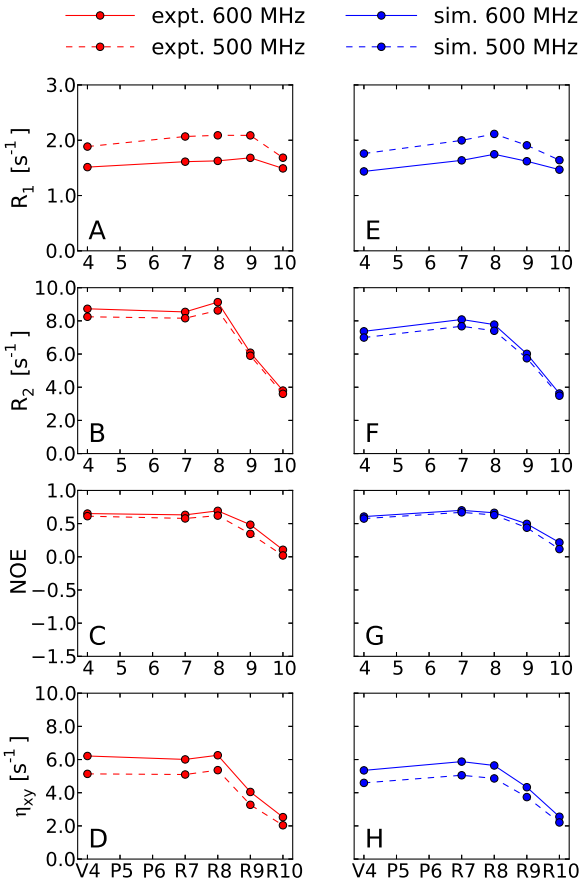


Figure 4.17. Experimental and simulated backbone ^{15}N relaxation rates in the Sos peptide bound to wtSH3 at 600 and 500 MHz spectrometer frequencies (solid and dashed lines, respectively). The simulated results are from the MD(wt-xray, $\lambda = 1.03$) trajectory.

at a certain distance from the surface of the protein ($\sim$ 15 Å considering the distance from the center of mass of Sos to the surface of wtSH3). The initial orientation of Sos was assigned randomly. The resulting system was used to construct a water box, allowing for a generous amount of water. Illustrated in Figure 4.18A is the initial setup that has been used to record one of the MD(rand, $\lambda = 1.03$) trajectories. Along with the single simulated wtSH3 molecule, the graph also shows three periodic images of wtSH3. This representation makes it possible to appreciate the amount of water in the system (the empty space in the graph is filled with TIP3P water). In this

particular trajectory, the effective concentration of the protein is 7.8 mM, which is not too far from the experimentally accessible concentrations (up to 4 mM in our experimental measurements, see Table 4.1). Consequently, there is enough room in this simulation for the Sos peptide to fully explore the conformational space available to it, without being squeezed by the protein molecules.

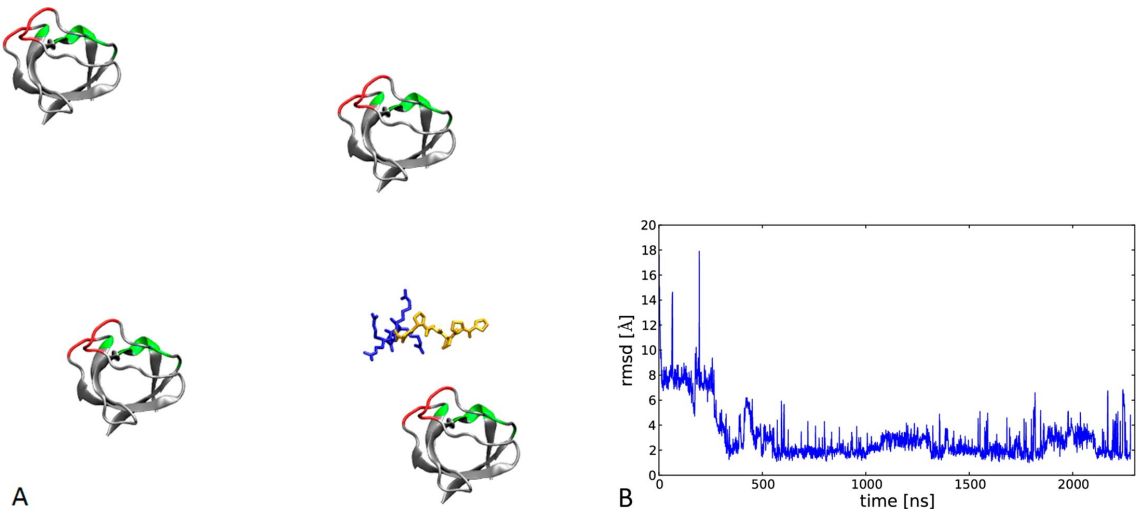


Figure 4.18. (A) Initial placement of the peptide and the protein in the 2.28 μs MD(wt-rand, $\lambda = 1.03$) trajectory. (B) rmsd of the peptide atomic coordinates relative to the crystallographic structure 1CKB.

Using this type of procedure, we have recorded thirty MD trajectories in order to model the binding process of Sos to wtSH3, and those MD trajectories have length ranging from 20 to 265 ns. In nine of these trajectories the Sos peptide and wtSH3 formed a complex that was found to be very similar, although not identical, to the crystal structure 1CKB. Of these nine trajectories, we have selected one that was subsequently extended to 2.28 μs. This trajectory, termed MD(wt-rand, $\lambda = 1.03$), was used to simulate the NMR observables, which were subsequently compared with the experimental values.

In the starting frame of the MD(wt-rand, $\lambda = 1.03$) trajectory, the Sos peptide is sufficiently far removed from the surface of wtSH3 (see Figure 4.18). Initially, the displacement of the Sos peptide (i.e., rmsd relative to the coordinates of the bound Sos as found in the crystal structure 1CKB) amounts to 17.6 Å. As the Sos·wtSH3 encounter complex is formed, the rmsd drops rapidly to $\sim$ 8 Å. The first salt bridge, R8–E167, is observed already after 4 ns. However, this turns out to be a transient interaction that immediately dissipates and subsequently makes only a few brief appearances throughout the trajectory. It is not before 45 ns that the two salt bridges are established in a sufficiently stable manner, R8–D150 and R8–D147. Both of these salt bridges are non-native. The resulting conformational state deserves a brief discussion. The pose of the peptide at this point generally resembles the binding pose found in 1CKB. In particular, the N-terminal portion of Sos hovers over the hydrophobic grooves where it is supposed to bind. However, the conformation of the peptide is not conducive to binding (there is a certain amount of twist): the peptide lies high above the surface of the protein and fails to form two signature hydrogen bonds, P2–Y186 and P5–W169. The hydrophobic buried surface area remains relatively small, and the peptide occasionally swings away from the surface of wtSH3, which causes the large spikes in the initial portion of the graph in Figure 4.18B.

After some time, this labile state is transformed into a *bona fide* complex. At around 270 ns, several conformational transitions take place, proline rings P2 and P5 slide into the shallow hydrophobic grooves on the surface of wtSH3, and the native hydrogen bonds P2–Y186 and P5–W169 are established. At around 290 ns, the (weak) native salt bridge R7–E149 is formed. At 320 ns, the non-native salt bridge R8–D150 is dissolved and replaced with the native salt bridge R7–D150. As a result of these transformations, by 340 ns the coordinate rmsd of the Sos peptide drops to 1.3 Å. However, large fluctuations continue to occur until 550 ns when the complex is stabilized by the hydrogen bond between the R8 amide group and the E166 side chain, as well as certain additional salt bridges. After that, the peptide mostly

remains within 2 Å of the crystallographic structure, although it still experiences substantial fluctuations involving the rearrangements of the C-terminal tail.

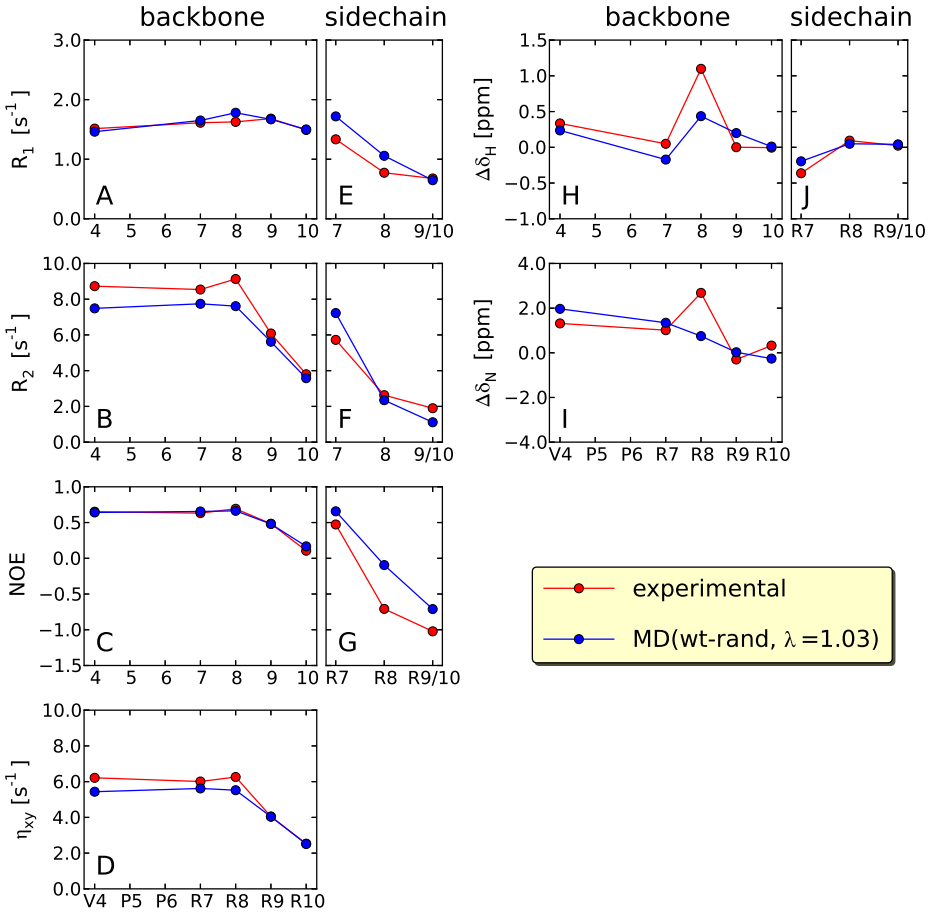


Figure 4.19. Comparison of the experimental (red) and simulated (blue) ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for the Sos peptide in complex with wtSH3 based on MD(wt-rand, $\lambda = 1.03$) trajectory. The 2.28 μs simulation MD(wt-rand, $\lambda = 1.03$) starts from the random placement of the peptide, and it uses the modified version of the Amber ff99SB*-ILDN force field in which the strength of salt bridge interactions involving Arg and Glu/Asp side chains has been adjusted.

The MD(wt-rand, $\lambda = 1.03$) trajectory has also been utilized to calculate ^{15}N relaxation parameters and binding shifts $\Delta\delta_{\mathrm{Sos}}$ to compare with experimental data. For

this purpose, we have chosen the portion of the trajectory that begins at 400 ns (the point at which the Sos:wtSH3 complex is fully formed). The results are presented in Figure 4.19. Briefly, they are virtually identical to those shown in Figure 4.15 with one exception—backbone shifts of residue R8 are significantly smaller than those observed experimentally (and those predicted based on the MD(wt-xray, $\lambda = 1.03$) trajectory). This deviation is caused by effective weakening of the hydrogen bond between R8 and the side chain of E166. In the selected time interval from 400 to 2280 ns, this hydrogen bond is identified in only 35% of all frames. By comparison, in the MD(wt-xray, $\lambda = 1.03$) trajectory, this proportion reaches 50%. Most likely, the difference can be attributed to the less-than-perfect convergence of the MD analyses. As already indicated, 2–3 μs simulations are sufficient to reliably reproduce ^{15}N relaxation rates, but not necessarily the binding shifts. However, in principle, we cannot rule out another possibility. It is conceivable that formation of the Sos:wtSH3 complex has not been fully completed in the 2.28 μs MD(wt-rand, $\lambda = 1.03$) simulation. Indeed, there are certain systematic differences between MD(wt-rand, $\lambda = 1.03$) and MD(wt-xray, $\lambda = 1.03$) trajectories—for example, the former features three intermittent salt bridges involving D147 that are virtually absent in the latter.

Finally, it is worthwhile to discuss the characteristic time scales as seen in the MD(rand, $\lambda = 1.03$) simulation. Using the values of k_{ON} and k_{OFF} experimentally determined in this study and the effective concentration of wtSH3 in the MD simulation, it can be estimated that the Sos:wtSH3 complex should be formed with a time constant of ~ 100 ns and dissolve with a time constant of ~ 500 μs. These estimates are consistent with the actual simulation, in which the final complex is formed after ~ 400 ns and the peptide never becomes separated from the protein over the entire duration of the MD trajectory. Note that we have been successful in observing the peptide binding only due to electrostatic interactions—diffusion-limited binding is much slower and thus most likely cannot be reproduced in the relatively short MD simulations.

4.4.2 Results about Sos·dmSH3 Electrostatic Encounter Eomplex

The MD simulations for Sos·dmSH3 electrostatic encounter complex have been carried out with both original and modified versions of Amber ff99SB*-ILDN force fields. The main results are based on the following three MD trajectories:

(i) The 6.0 μs trajectory of Sos:dmSH3 starting from the crystallographic coordinates 1CKB recorded with the standard Amber ff99SB*-ILDN force field, $\lambda = 1.00$. This trajectory is termed MD(dm-xray, $\lambda = 1.00$).

(ii) The 6.0 μs trajectory of Sos:dmSH3 starting from the crystallographic coordinates 1CKB recorded with the altered Amber ff99SB*-ILDN force field, $\lambda = 1.03$. This trajectory is termed MD(dm-xray, $\lambda = 1.03$).

(iii) The 6.0 μs trajectory that represents the binding of Sos to dmSH3 starting from a random configuration. This trajectory has been recorded using the altered Amber ff99SB*-ILDN force field, $\lambda = 1.03$, and is termed MD(dm-rand, $\lambda = 1.03$).

4.4.2.1 Simulation of Sos·dmSH3 using Amber ff99SB*-ILDN

The MD simulation for Sos·dmSH3 complex has been carried out with the original Amber ff99SB*-ILDN force field, still using X-ray structure (PDB 1CKB) as the initial conformation. All parameters used for MD simulations such as the choice of water model and the size of water box were the same as those used for recording MD(wt-xray, $\lambda = 1.00$) trajectory. This MD trajectory has been extended to 6.0 μs. ^{15}N relaxation parameters and binding shifts of Sos peptide $\Delta\delta_{\mathrm{Sos}}$ have been calculated from this MD trajectory to compare with experimental data.

The center-of-mass distribution of Sos peptide around dmSH3 generated from this MD trajectory is shown in Figure 4.20. By comparing center-of-mass distributions obtained from Sos·dmSH3 and Sos:wtSH3 (Figures 4.20 and 4.10), it is quite clear that Sos peptide within Sos·dmSH3 complex should behave much more flexible than

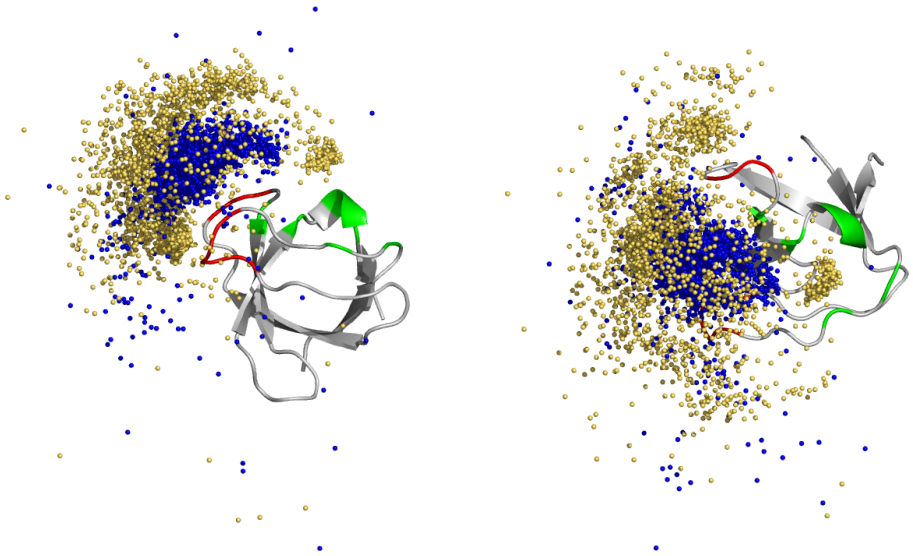


Figure 4.20. Position of the peptide Sos on the surface of dmSH3 according to the data from the 6.0 μs simulation MD(dm-xray, $\lambda = 1.00$). The plotting conventions are the same as used in Figure 4.10, except for using 2 ns sampling interval for MD frames.

within Sos:wtSH3 complex. The N-terminal portion of Sos peptide frequently jumps outside the binding pocket of dmSH3, which is mainly due to weakened binding affinity caused by introduced mutations, and such behavior has never been observed on MD trajectories from Sos:wtSH3 complex. It can also be recognized that Sos peptide still stays around the original binding interface most of the time, it only occasionally shows up at some other sites around the surface of dmSH3.

The comparison between MD simulation and NMR experimental results are shown in Figure 4.21, the agreement between the two datasets is not satisfactory. The main problem is similar to the case of simulating Sos:wtSH3—several residues show much elevated ^{15}N relaxation parameters (such as R_2, η_{xy} and NOE) than experimental data, which strongly suggests that the motional freedom of Sos peptide is still too much restricted in this MD trajectory. In previous studies about Sos:wtSH3, the problem caused by overestimated electrostatic interactions (i.e., salt bridges) has

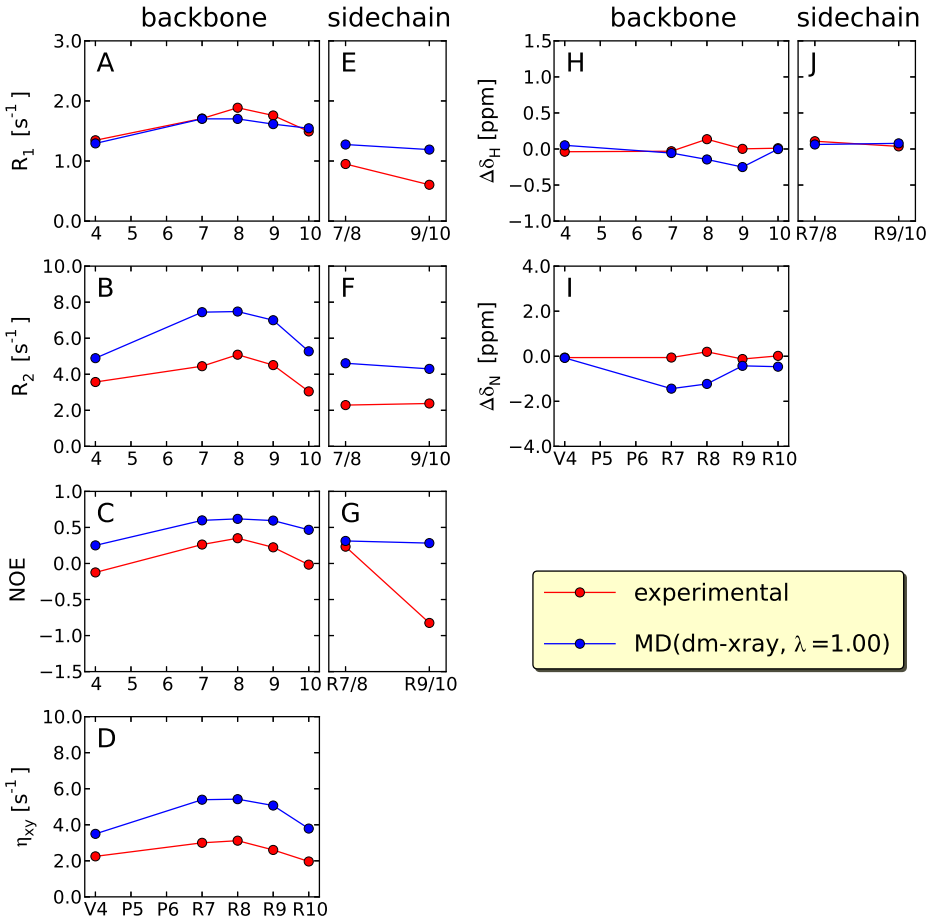


Figure 4.21. Comparison of the experimental (red) and simulated (blue) ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for the Sos peptide in complex with dmSH3 based on MD(dm-xray, $\lambda = 1.00$) trajectory. Simulation MD(dm-xray, $\lambda = 1.00$) uses the original Amber ff99SB*-ILDN force field and starts from the crystallographic coordinates 1CKB.

been resolved by applying modified Amber ff99SB*-ILDN force field, therefore such strategy could also help to overcome the issue encountered for simulating Sos·dmSH3.

4.4.2.2 Simulation of Sos·dmSH3 Using Modified Version of Amber ff99SB*-ILDN

The MD simulation for Sos·dmSH3 complex with modified Amber ff99SB*-ILDN force field has been carried out with X-ray structure as initial conformation. The modified version of Amber force field still applies $\lambda = 1.03$ to adjust LJ parameters

(i.e., the equilibrium distance σ) on specific atom pairs based on results from studying Sos:wtSH3 complex. All other parameters are the same as used for recording MD(dm-xray, $\lambda = 1.00$) trajectory, this MD trajectory MD(dm-xray, $\lambda = 1.03$) has also been extended to 6.0 μs.

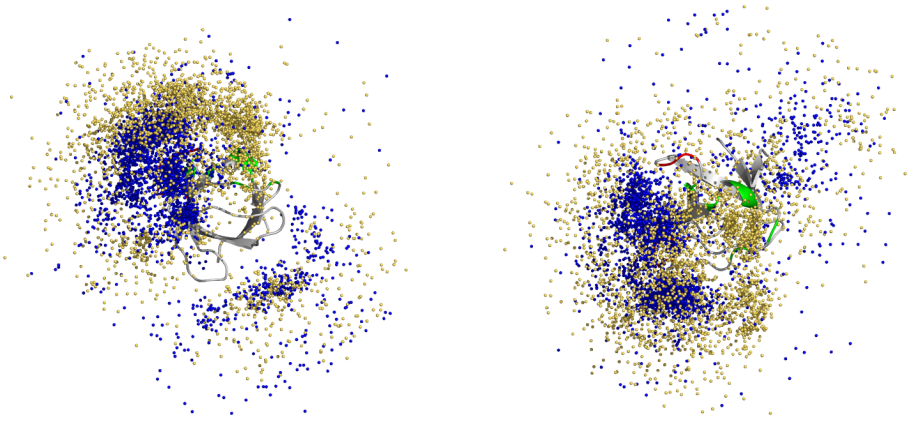


Figure 4.22. Position of the peptide Sos on the surface of dmSH3 according to the data from the 6.0 μs simulation MD(dm-xray, $\lambda = 1.03$). The plotting conventions are the same as used in Figure 4.20.

The behavior of Sos peptide in this MD trajectory can be recognized from the center-of-mass distribution plot (Figure 4.22). By comparing Figure 4.22 and Figure 4.20, it is easy to see that Sos peptide in MD(dm-xray, $\lambda = 1.03$) trajectory behaves much more flexible than in MD(dm-xray, $\lambda = 1.00$) trajectory—both N- and C-terminal portions produce even larger center-of-mass distributions, besides, Sos peptide also spends more time on some other binding sites other than the original X-ray binding site, therefore this encounter complex exists as a heterogeneous conformational ensemble. Such structural features have also been observed on some other well-studied encounter complexes, such as cytochrome c and cytochrome c peroxidase[145]. The large mobility of Sos peptide within electrostatic encounter complex is necessary to ensure that it can perform complete conformational search in order to

find the correct final binding site. Figure 4.22 also indicates that the center-of-mass of Sos peptide still locates near the X-ray binding site most of the time, therefore the conformational search is not completely random, but it still prefers those conformations that are more likely to form the correct final binding pose. This feature also agrees with results as obtained from the reverse titration experiment, which also suggests that Sos peptide interacts with dmSH3 near the X-ray binding site mostly.

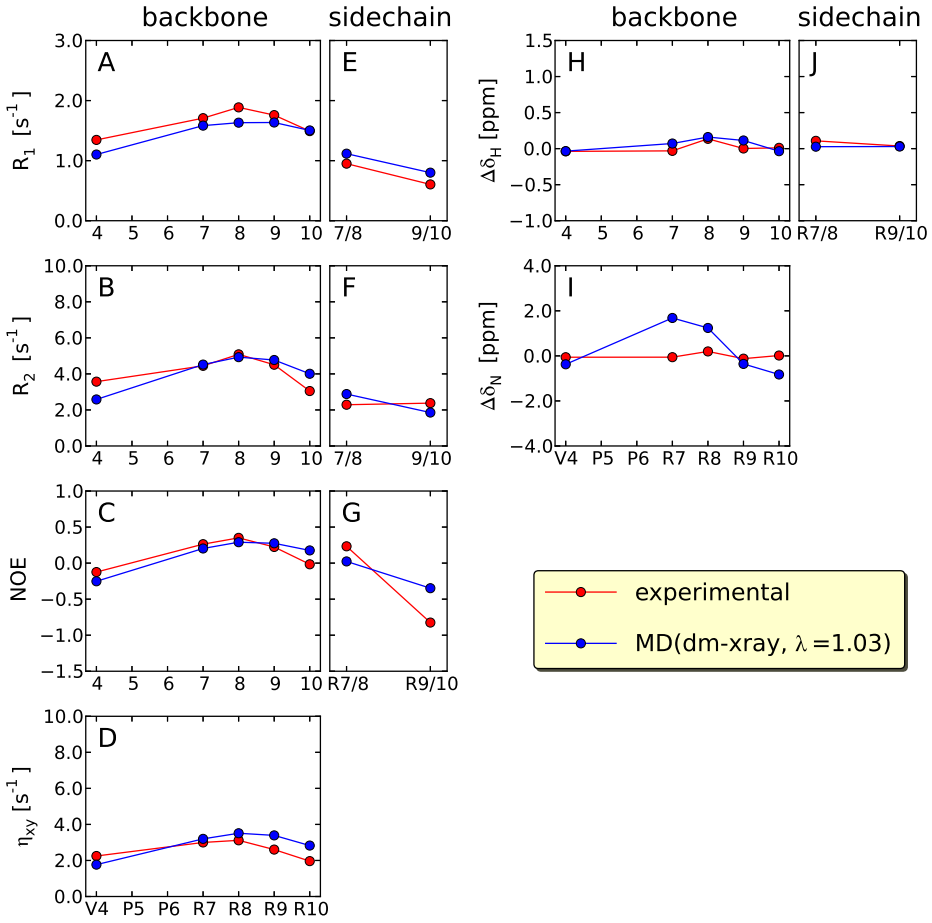


Figure 4.23. Comparison of the experimental (red) and simulated (blue) ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for the Sos peptide in complex with dmSH3 based on MD(dm-xray, $\lambda = 1.03$) trajectory.

The comparison of ^{15}N relaxation parameters and binding shifts $\Delta\delta_{\mathrm{Sos}}$ for Sos peptide between MD simulation and experimental results are shown in Figure 4.23. The agreement is much better than obtained from the original Amber force field (Figure 4.21), specifically, ^{15}N relaxation parameters such as R_2, η_{xy} and NOE are much closer to experimental data, which is mainly due to the larger mobility of Sos peptide within this MD trajectory. Since MD(dm-xray, $\lambda = 1.00$) and MD(dm-xray, $\lambda = 1.03$) only differ on LJ parameters for specific atomic pairs within Arg–Asp/Glu residue pairs, it indicates that those LJ parameters are crucial for correctly modeling the behavior of Sos·dmSH3 encounter complex, although such modification is quite subtle (i.e., from $\lambda = 1.00$ to $\lambda = 1.03$). The fact that modified Amber ff99SB*-ILDN force field ($\lambda = 1.03$) produces better agreement with experimental data for both Sos:wtSH3 and Sos·dmSH3 complexes indicates that such modification is likely to be universal and it should also be applied for simulating other similar systems. However, the determination of more accurate rescaling parameter λ may still require further rounds of MD force field optimizations, based on more systematic force field refinement strategy.

During the MD(dm-xray, $\lambda = 1.03$) trajectory, Sos peptide frequently jumps between different binding sites on dmSH3 (sometimes it flies away from dmSH3), which induces the large conformational heterogeneity of this complex. Such feature basically agrees with experimentally determined k_{off} ($\sim 3 \times 10^5$ s^{-1}), which indicates that the average lifetime of Sos·dmSH3 complex is ~ 3 μs, therefore it is possible to observe multiple dissociation events from $\sim\mu$s timescale MD simulations, and this cannot be achieved for Sos:wtSH3 due to the much lower k_{off}. Whenever Sos peptide dissociates from dmSH3, it always quickly re-associates with the original copy or another periodic image of dmSH3 due to strong electrostatic attraction forces. Therefore during the whole length of 6.0 μs MD(dm-xray, $\lambda = 1.03$) trajectory, Sos peptide is in un-

bound state only for a very small fraction of time ($< 1\%$), such that the original ^{15}N relaxation parameter calculation strategy for Sos peptide should still remain valid.

4.4.2.3 Simulation of Sos·dmSH3 with Different Initial Conformations

Although MD simulation results from MD(dm-xray, $\lambda = 1.03$) trajectory are in good agreement with experimental data, since Sos peptide displays large amount of motional freedom, it is worth considering the effects of using different initial conformations. The most ideal case is that the same results can always be reproduced and are close to experimental data, such that the correct behavior of encounter complexes can always be captured with any choice of initial conformation. In order to investigate that we have recorded another ten MD trajectories with Sos peptide randomly placed around dmSH3. It turns out that the behavior of Sos peptide as well as simulated ^{15}N relaxation parameters show large variations among different MD trajectories, even though each MD trajectory has been extended to at least ~ 1 μs. In certain MD trajectories, Sos peptide behaves too rigid and get trapped at one single binding site, therefore it cannot produce ^{15}N relaxation parameters in good agreement with experimental data. In some other MD trajectories, Sos peptide dissociates from and re-associates with dmSH3 too frequently, therefore it displays too much flexibility such that cannot produce results close to experiments either. Among all MD trajectories have been recorded, only two or three show similar properties with MD(dm-xray, $\lambda = 1.03$) trajectory. One of those MD trajectories has been extended to 6.0 μs for further analysis, which is termed MD(dm-rand, $\lambda = 1.03$).

The center-of-mass distribution of Sos peptide around dmSH3 as obtained from MD(dm-rand, $\lambda = 1.03$) trajectory is shown in Figure 4.24, which displays much similar property as from MD(dm-xray, $\lambda = 1.03$) trajectory (Figure 4.22): the center-of-mass distribution of Sos peptide covers large fraction of dmSH3 surface area, and still those positions near the X-ray binding site are more preferred. Besides, ^{15}N relaxation parameters and binding shifts of Sos peptide $\Delta\delta_{\mathrm{Sos}}$ calculated from this MD trajectory also show good agreement with experimental data (Figure 4.25).

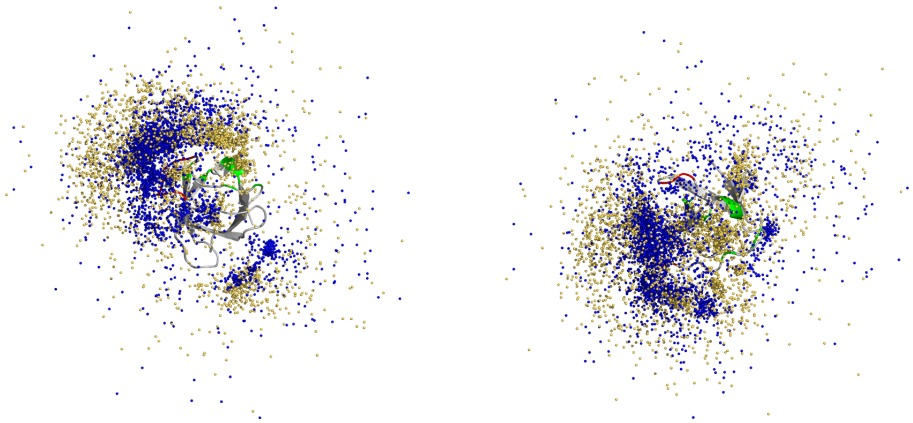


Figure 4.24. Position of the peptide Sos on the surface of dmSH3 according to the data from the 6.0 μs simulation MD(dm-rand, $\lambda = 1.03$). The plotting conventions are the same as used in Figure 4.20.

Even though the properties of Sos·dmSH3 encounter complex have been well captured by both MD(dm-xray, $\lambda = 1.03$) and MD(dm-rand, $\lambda = 1.03$) trajectories, based on results from all the MD trajectories have been recorded, it should still be concluded that the convergence issue still exists and there is no guarantee that the behaviors of encounter complex can always be correctly modeled with any choice of initial conformation. The convergence issue also commonly exists for many other MD simulation studies, the fact that distinct properties are observed from different MD trajectories usually indicates that different regions of the conformational space have been sampled. In the case of Sos·dmSH3 encounter complex, since Sos peptide displays large amount of motional freedom, the complete conformational space corresponding to this complex should be quite large (at least much larger than Sos:wtSH3), therefore it is still difficult to achieve complete conformational sampling even with $\sim\mu$s timescale MD simulations. Although different types of advanced conformational sampling methods have been developed for overcoming such issues, they usually produce MD trajectories that are not realistic such that cannot be used to calculate experimental parameters such as ^{15}N relaxation data. It can be expected that it should be possible to achieve near-complete conformational sampling based on one

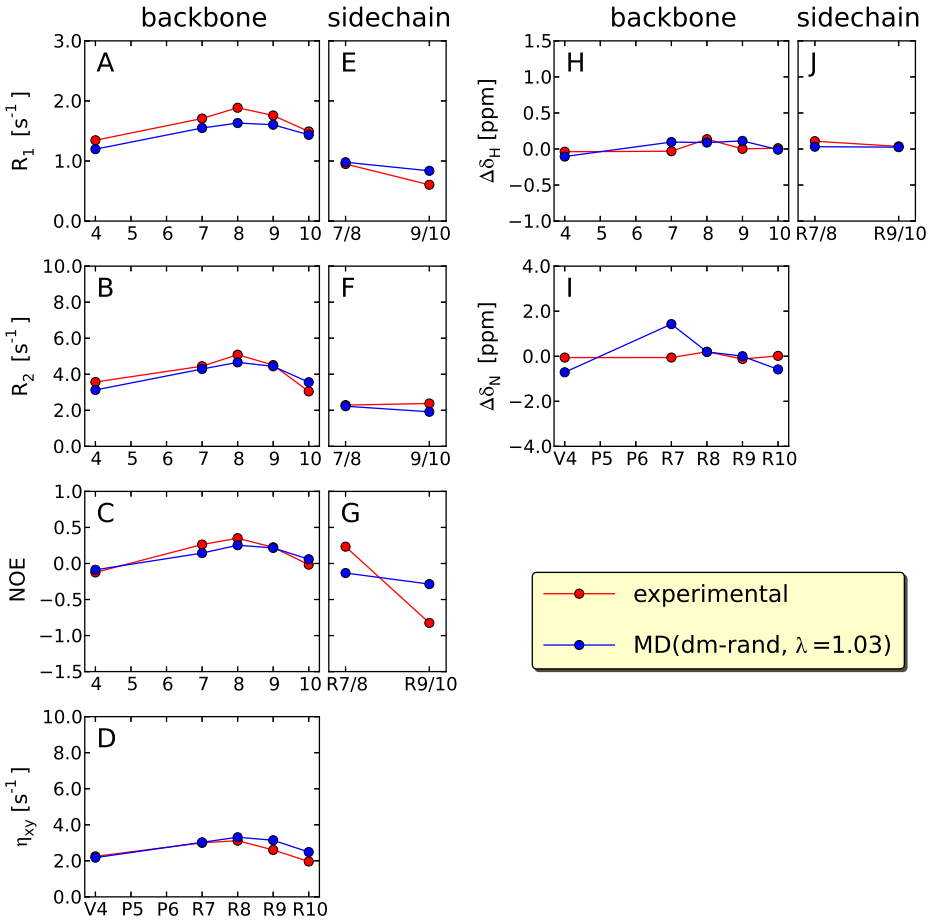


Figure 4.25. Comparison of the experimental (red) and simulated (blue) ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for the Sos peptide in complex with dmSH3 based on MD(dm-rand, $\lambda = 1.03$) trajectory.

ultralong MD trajectory, although nowadays it is still difficult to record ~ms timescale MD trajectories in a routine fashion, in future this might be possible with even better computer hardwares and MD programs. There is still another issue concerning about the quality of MD force field itself: one important lesson has been learned from this study is that once MD trajectory has reached very long, even a small imperfection in MD force field may results in large discrepancy with experimental data. Therefore it

is still essential to further optimize most existing MD force fields and improve their qualities.

4.5 Concluding Remarks

MD modeling combined with experimental NMR studies has been enormously popular in the context of (modestly sized) globular proteins. Such combined studies produced a wealth of information, greatly advancing our understanding of protein dynamics and at the same time stimulating the development of force fields. However, much of local dynamics in globular proteins is somewhat generic: for example, small-amplitude fluctuations of peptide planes can be accurately predicted on the basis of protein structure, thus to some degree obviating the need in specialized MD studies. In contrast, intrinsically disordered proteins display a rich range of dynamic behaviors that are uniquely suited for MD and NMR studies. Yet until recently the progress in this area has been impeded by a lack of adequate computational resources. Indeed, proper modeling of an IDP requires a very large water box to accommodate the more extended protein conformations; it also requires a very long trajectory to sample the vast conformational space available to an IDP. All of this has become possible only with the advent of new computers and the suitably adapted MD algorithms.

In our study, we investigate the protein–peptide system that features a substantial element of dynamic disorder. For example, in the MD(wt-rand, $\lambda = 1.03$) simulation, we model the flexible 10-residue peptide, which moves in a stochastic fashion towards its protein target and then forms a highly dynamic intermediate before making a transition to the *bona fide* complex. Furthermore, even in the bound state, the peptide retains a substantial amount of motional freedom, especially the C-terminal residues. Although the presence of disorder in this system clearly raises the bar for MD studies (in terms of both the simulation length and the size of the water box), the simulation has been successful, reproducing the experimental NMR data with near-quantitative accuracy. We envisage that in the near future this methodology will progress from peptides to small disordered proteins and eventually to large IDPs,

improving our understanding of many aspects of their function, including the disorder-to-order transition upon binding to their folded targets.

In the first part of this study, we focus on the moderately high affinity ($\sim 1\ \mu$M) complex between Sos and wild-type SH3. Comparing the experimental data (specifically, ^{15}N relaxation rates) with the results from the state-of-the-art Amber ff99SB*-ILDN simulation, we have found that molecular dynamics overemphasizes salt bridge interactions between the peptide and the protein. The tendency to misjudge the strength of salt bridges appears to be a common affliction of modern fixed-charge force fields.

To address this problem, we have devised a targeted correction involving the Lennard–Jones potential between the charged moieties of the Arg and Asp/Glu side chains. This is clearly an empirical solution: the proper way of approaching this problem would be to systematically reparameterize the entire force field. Nevertheless, there is little doubt that off-diagonal Lennard–Jones parameters associated with salt bridges need to be reassessed in Amber ff99SB as well as many other widely used MD force fields. In this sense, we have likely identified an important target for future optimization of non-polarizable force fields.

The MD trajectory of Sos:wtSH3 recorded with the altered force field, Amber ff99SB*-ILDN ($\lambda = 1.03$), showed good agreement with the experiment, specifically, with respect to ^{15}N relaxation rates and binding shifts (in the latter case, the outcome is to some degree affected by convergence). As it turns out, weakening of the salt bridges in the C-terminal portion of Sos strengthens the binding at the N-terminal region. This implies that there is a certain amount of conflict between, on one hand, the salt bridge interactions implicated in formation of the electrostatic encounter complex and, on the other hand, hydrogen bonds and hydrophobic packing indicative of the tight complex. In other words, the interactions associated with the electrostatic encounter complex do not necessarily constitute a subset of the interactions found in the fully formed complex. As it appears, the weakening of the salt bridges in the $\lambda = 1.03$ simulation has an effect not so much on the amplitudes of motion of the

C-terminal residues but on the respective correlation times. Compared to the original simulation, the C-terminal tail visits the same conformations but does not remain in these conformations for quite as long. Faster conformational exchange is reflected in the spin relaxation rates, leading to improved agreement with the experiment.

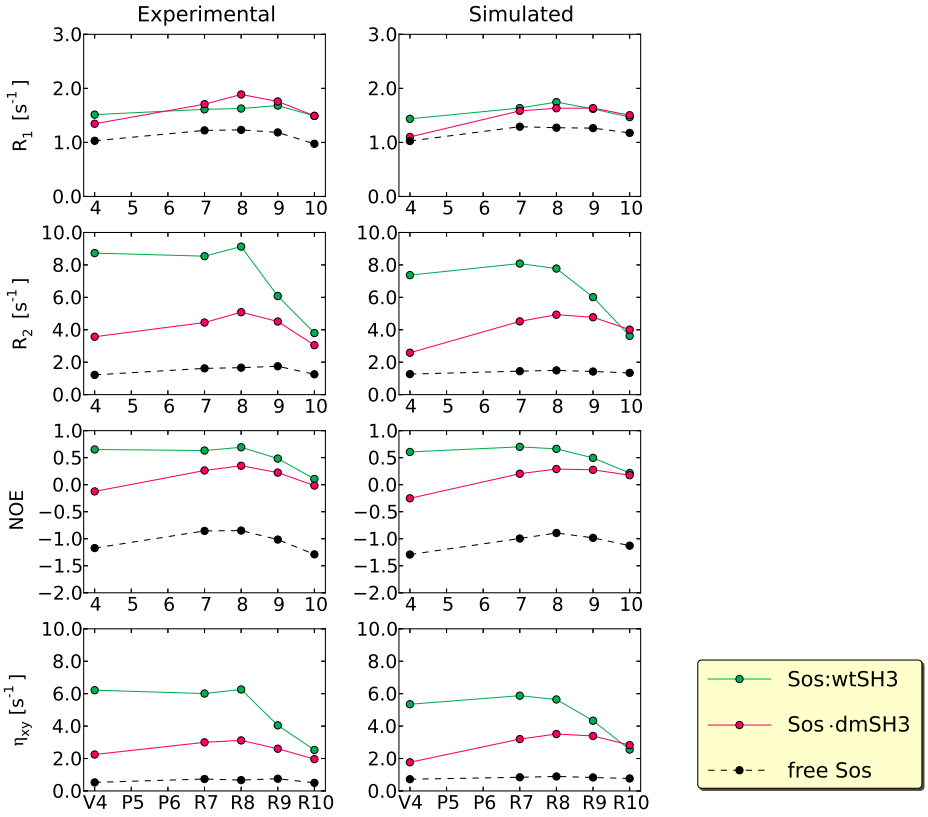


Figure 4.26. Comparison of the experimental and simulated ^{15}N relaxation parameters for the Sos peptide in three different forms. The simulated results are all based on the modified version of the Amber ff99SB*-ILDN force field in which the strength of salt bridge interactions involving Arg and Glu/Asp side chains has been adjusted (i.e., $\lambda = 1.03$). The MD simulation results for Sos:wtSH3 and Sos·dmSH3 are both based on the MD trajectories with X-ray initial structure.

In the second part of this study, we use the same approach to directly characterize the system that mimics the electrostatic encounter complex. Specifically, we have

designed the double mutant of c-Crk N-SH3 (Y186L/W169F) in which the binding of the N-terminal portion of Sos is largely abrogated and the complex is held together mainly by the salt bridges, thus resembling the electrostatic encounter complex. The electrostatic encounter complex Sos·dmSH3 has also been successfully characterized via NMR measurements and modeled with the help of the (modified) MD protocol. The relaxation parameters for Sos peptide in three different forms are summarized in Figure 4.26. It is quite clear that the three different dynamic states of Sos peptide can be well distinguished based on ^{15}N relaxation parameters, besides, our MD simulation results based on modified Amber force field well reproduce all the experimental data. Such combined MD and NMR approach can be utilized as a general strategy to investigate structural and dynamic properties of many other electrostatic encounter complexes that are involved with disordered proteins or peptides.

4.6 Materials and Methods

4.6.1 NMR Sample Preparation

4.6.1.1 Protein and Peptide Expression

A primer for producing c-Crk N-SH3 was designed in house using mouse cDNA (GenBank accession number BC031149) and GST plasmid pGEX-4T-1. For the wild-type protein wtSH3, the original cDNA was used without any modification. For W169F/Y186L double-mutant dmSH3, it was constructed based on wtSH3 using QuikChange Site-Directed Mutagenesis Kit. The protein was expressed and purified as described previously[497] to produce unlabeled, ^{15}N-labeled, or ^{15}N- and ^{13}C-labeled material. The ^{15}N-labeled Sos peptide was bacterially expressed as a fusion construct with a solubility tag protein, the GB1 domain from protein G[498,499]. The Sos sequence was subcloned into GB1-containing plasmid GEV 2, with a single spacer residue (Asp). Following the carefully optimized expression and purification procedure, GB1-Sos was cleaved at the Asp-Pro site using a treatment with a 50% (v/v) solution of

formic acid (24 h at 50 °C). The products were separated using reverse phase HPLC with C18 column. The sample was concentrated using a lyophilizer or SpeedVac apparatus; the buffer exchange was conducted over a period of at least 24 h using a dialyzer device with a 0.5 kDa membrane cutoff. The yield from this procedure was 1–2 mg of ^{15}N-labeled Sos from 1 L of M9 medium. Unlabeled Sos, as well as Sos selectively labeled with ^{15}N at the V4 position, was purchased commercially.

4.6.1.2 Sample Conditions

Unless otherwise indicated, most samples used in this work have been prepared in a buffer containing 90% H_2O/10% D_2O, 20 mM phosphate, and 0.02% NaN_3 (pH 6.0). The concentrations of the peptide and protein were determined using NMR-based method[500], which ensures an accuracy on the order of 3%. Specifically, the intensities of isolated methyl resonances (V4 in Sos and V184 in c-Crk N-SH3) from one-dimensional (1D) proton spectra were determined first then compared with the external reference (10 mM sample of valine). For c-Crk N-SH3, the results were consistent with the standard approach based on UV absorption by aromatic residues, A_{280}. All NMR data were collected at 23.4 °C.

4.6.2 NMR Experiments

4.6.2.1 NMR Measurement

NMR measurements were conducted using a Varian Inova 600 MHz spectrometer equipped with an HCN probe; additionally, ^{15}N relaxation parameters were measured using a Bruker Avance DRX 500 MHz spectrometer equipped with a TXI cryoprobe. The backbone ^{15}N relaxation measurements were conducted using updated versions of the standard relaxation experiments[135,338,501], including the corrected ^{15}N–{^{1}H} NOE sequence[304]. ^{15}N DD/CSA transverse cross-correlation rate η_{xy} has been measured by means of the IPAP sequence[60] with an added relaxation period with a length that is a multiple of $1/^1J_{NH}$[450]. For arginine side-chain $^{15}N^{\varepsilon}$ relaxation measurements, 1D

versions of the same experiments were used, with the nitrogen *rf* carrier set to 80 ppm. The data have been processed using the NMRPipe suite of programs[261], including the *autofit* script, and Sparky. The three-dimensional HNCACB[502] experiment was carried out for ^{15}N-,^{13}C-labeled dmSH3 sample to confirm the assignment of dmSH3.

4.6.2.2 Spectral Assignments

The resonance assignments for wtSH3 was already available before this work[503]. The resonance assignments for dmSH3 is largely based on the assignment of wtSH3, since most peaks in ^{1}H–^{15}N HSQC spectra of these two proteins show good overlap, except for those residues close to the mutation sites of W169F and Y186L. The assignment for dmSH3 is based on results from the HNCACB experiment, which correlates H^{N} with C^{α}/C^{β} atoms in the same and the previous amino acid residues[502]. This experiment alone can already provide unambiguous assignment for dmSH3, therefore there is no need to perform any additional assignment experiments (such as HN(CO)CACB).

The resonance assignment for backbone amides in Sos is obtained from standard NOESY and TOCSY experiments performed on a 10 mM sample of unlabeled Sos[504]. The assignment has been confirmed by solvent exchange data: at pH 7.0, the resonances from R8 and R9 are broadened beyond detection, which is consistent with expectations[350]. They were also confirmed using a home-written triple-resonance experiment correlating $^{1}H^{\varepsilon}$,$^{15}N^{\varepsilon}$ spins with $^{1}H^{N}$,^{15}N spins of the subsequent residue. The spectrum of the free peptide also contains a number of minor peaks arising from the *cis* conformation of adjacent proline residues[505]; some of these minor peaks have also been assigned. Four arginine $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ correlations in the spectrum of the free Sos peptide are overlapped, resulting in a single unresolved peak. Two of these resonances become resolved and can be assigned in the complex with wtSH3. In the complex with dmSH3 it produces two groups of peaks with similar intensity. $^{1}H^{\eta}$–$^{15}N^{\eta}$ correlations were found to be severely broadened and have not been used.

4.6.3 MD Smulations

4.6.3.1 MD Simulation Setup

The main series of MD simulations has been conducted using the Amber 11 package with the AMBER ff99SB force field[201,451], including ILDN side-chain correction[203] and backbone helical propensity correction[202]. The initial coordinates of the system were generated using two different approaches.

In the first approach, the initial coordinates were based on the crystallographic structure 1CKB[197]. The missing Sos residues R9 and R10 were added with random conformation using the program MODELLER[456]. The SH3 domain was simulated without three terminal residues, G132, S133, and R191, which were present in our NMR sample. We also conducted control simulations, in which these three residues were added to the crystallographic structure; the results proved to be unchanged.

In the second approach, the peptide was placed over the surface of the protein in a random fashion. Specifically, in the crystal structure 1CKB, we defined a vector connecting the centers of mass of SH3 and Sos. The length of this vector was doubled from 12.6 to 25.2 Å, and the obtained point was used to place the center of mass of the Sos peptide. The initial conformation of peptide was chosen at random from a 2.1 μs MD simulation of the free Sos. The initial orientation of Sos was assigned randomly. Finally, a random rotation was applied to SH3 (the amplitude of this rotation was normally distributed with a variance of 30°).

All Glu and Asp residues in the SH3 domain were assumed to be deprotonated under the conditions of our experimental study (pH 6.0). This is consistent with the calculation employing the program PROPKA[251] and the crystallographic structure 1CKB. These calculated pK_a values are all sufficiently low (the highest value, 5.0, is predicted for residue E149). To verify this result, we have prepared a sample of Sos:wtSH3 with an elevated pH (7.0) where it is safe to assume that all Asp and Glu side chains are deprotonated. We have found that the chemical shifts of Sos, including $^{1}H^{\eta}$ and $^{15}N^{\eta}$ shifts, remain essentially unchanged in this sample. Therefore,

we can conclude that Asp and Glu residues at the Sos binding interface are already deprotonated at pH 6.0.

The protein–peptide system thus generated was solvated using the explicit TIP3P solvent. A truncated octahedral water box was constructed such that the boundary of the water box was at least 12 Å away from any of the peptide or protein atoms. Given that the peptide is initially placed at considerable distance from the protein, the resulting water box contains a generous amount of water. As a result, the effective concentration of the protein in our simulations was relatively modest, 6–8 mM. This is not too far from the experimental NMR concentration. The above description mainly pertains to the setup with random placement of Sos. In the case when the initial configuration was based on 1CKB, the water box with a 14 Å margin was used, however, because the 1CKB complex is sufficiently compact, the water box proved to be even smaller than those with random placement of Sos, corresponding to an effective protein concentration 13–15 mM.

The hydrated system was neutralized by adding one Na^+ ion (for the protein missing R191) and subjected to energy minimization (1000 steps using harmonic restraints with a force constant of 500 kcal $mol^{-1}Å^{-2}$, followed by 1000 steps with no restraints), then heated from 0 to 296 K, and equilibrated for 1 ns at 296 K. The production-stage MD simulations were conducted at 296 K using the NPT ensemble. During the simulations, all bonds involving hydrogen atoms were constrained using the SHAKE algorithm. The nonbonded cutoff was set to 8 Å. The integration step was 2 fs, and the coordinates were stored every 1 ps. The simulations were conducted using two GPU workstations, one equipped with four NVIDIA GeForce GTX480 cards and the other with four GTX580 cards. The production rate using the CUDA version of the pmemd program was in the range from 30 to 60 ns per day per card (depending on the size of the water box).

To adjust the strength of the salt bridge interactions in the Amber ff99SB*-ILDN force field, we have implemented one change in the original force field. The Lennard–

Jones energy for a pair of atoms separated by distance r is parametrized in Amber as follows:

$$E_{LJ} = \varepsilon \left[\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^{6} \right] \tag{4.4}$$

where ε is the well depth and r_0 is the equilibrium distance. We have altered the r_0 pertaining only to certain specific atom pairs, namely the nitrogen–oxygen pairs involving N^{η} atoms in Arg side chains and O^{ϵ}/O^{δ} atoms in Glu/Asp side chains. The original Amber-generated value of r_0 for these atomic pairs has been multiplied by a scaling factor λ. Because the GPU version of the Amber 11 pmemd program does not use look-up tables, this was accomplished by making a simple modification to the source code (kNLCPNE.h). This modification is not necessary for the latest version of Amber 14 pmemd program (GPU version), which also handles the nonbonded LJ energy term using look-up tables (similar to the CPU version of pmemd program).

4.6.3.2 MD Trajectories Processing and NMR Parameter Calculation

The MD trajectories were processed using in-house C++ package trjtool. The companion package pytrj, programmed in Python, serves to calculate and plot various relaxation parameters such as ^{15}N relaxation rates.

To calculate chemical shifts, we have extracted every 10th frame from a given trajectory, processed these frames using SHIFTX2 version 1.07[461], and then averaged the results. All computations were conducted using the SHIFTX+ module, which deals with the conformational dependence of chemical shifts. To investigate the effect of sampling, we repeated the calculations using all MD frames; the results were found to be virtually identical. Our focus is on the differential chemical shifts, $\Delta\delta_{\mathrm{Sos}} = \delta_{\mathrm{Sos}}^{\mathrm{bound}} - \delta_{\mathrm{Sos}}^{\mathrm{free}}$, which are termed binding shifts. In the calculations, the $\delta_{\mathrm{Sos}}^{\mathrm{bound}}$ component has been obtained from Sos:wtSH3 trajectories whereas $\delta_{\mathrm{Sos}}^{\mathrm{free}}$ has been obtained from the simulation of the free Sos. The effect of weak self-association of wtSH3 on $\Delta\delta_{\mathrm{Sos}}$ is inconsequential and has been ignored in these calculations.

To calculate the relaxation rates, all MD frames were aligned with the crystallographic coordinates 1CKB by superimposing the protein coordinates based on C^{α} atoms from the secondary structure regions. In this way only the protein coordinates have been superimposed but the motion of the peptide relative to the protein surface is fully preserved. The overall tumbling of the Sos:wtSH3 complex was then reintroduced using the experimentally determined correlation time, $\tau_R = 7.0$ ns, which already accounts for weak self-association effect determined based on wtSH3. The autocorrelation functions have been computed for NH vectors using a sparse nonuniform time grid as described previously[506]; they were subsequently fit with six-exponential functions, multiplied by $\exp(-\tau/\tau_{\mathrm{R}})$, and then converted into spectral densities and further into relaxation rates[507]. In the case of DD/CSA cross-correlated cross-relaxation rate η_{xy}, the result was multiplied by $P_2(\cos\theta)$ assuming that the long axis of the ^{15}N CSA tensor makes an angle $\theta = 20^{\circ}$ with the NH bond[508]. The length of the NH bond was assumed to be 1.02 Å, and the magnitude of the nitrogen chemical shift anisotropy was taken to be -172 ppm. More details about ^{15}N relaxation parameter calculation from MD trajectories are included in Appendix D.

The in-house Python script has been employed to analyze the appearance of salt bridges and hydrogen bonds in the MD trajectory. In the case of salt bridges, at least one nitrogen atom from the Arg side chain has been required to fall within 4 Å of the oxygen from the Glu/Asp side chain. In addition, the distance between the centroids of the two charged groups should not exceed 5 Å. This definition covers both high- and medium-strength ion–ion contacts, which mainly play a stabilizing role[509]. In the case of hydrogen bonds, we have required that the nitrogen–oxygen distance be < 3.2 Å and the deviation from linearity not exceed 30°[510].

The solvent-accessible surface area was calculated with per frame basis using the program POPS using the default parameter settings[511].

4.7 Major Contribution into This Thesis by Other Investigators

This work is equally contributed by Tairan Yuwen and Yi Xue. The idea of modifying LJ parameters benefits a lot from the discussion with Fangqiang Zhu. Lake Paul provided some help for the setup of ITC experiments. Greg Jarvis in Prof. Low's lab provided much help for Sos peptide purification with HPLC.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] C.B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 1973, 181(4096):223–230.

[2] J.H. Han, S. Batey, A.A. Nickson, S.A. Teichmann, and J. Clarke. The folding and evolution of multidomain proteins. *Nature Reviews Molecular Cell Biology*, 2007, 8(4):319–330.

[3] G. Cesareni, M. Gimona, M. Sudol, and M. Yaffe. *Modular Protein Domains*. Wiley-VCH, Weinheim, 2005.

[4] R.S. Gokhale and C. Khosla. Role of linkers in communication between protein modules. *Current Opinion in Chemical Biology*, 2000, 4(1):22–27.

[5] B.Y. Ma, C.J. Tsai, T. Haliloglu, and R. Nussinov. Dynamic allostery: Linkers are not merely flexible. *Structure*, 2011, 19(7):907–917.

[6] V.P.R. Chichili, V. Kumar, and J. Sivaraman. Linkers in the structural biology of protein–protein interactions. *Protein Science*, 2013, 22(2):153–167.

[7] X.Q. Wang, P. Vallurupalli, A. Vu, K. Lee, S. Sun, W.J. Bai, C. Wu, H.J. Zhou, J.E. Shea, L.E. Kay, and F.W. Dahlquist. The linker between the dimerization and catalytic domains of the CheA histidine kinase propagates changes in structure and dynamics that are important for enzymatic activity. *Biochemistry*, 2014, 53(5):855–861.

[8] J.J. Chou, S.P. Li, C.B. Klee, and A. Bax. Solution structure of Ca^{2+}–calmodulin reveals flexible hand-like properties of its domains. *Nature Structural Biology*, 2001, 8(11):990–997.

[9] Y.J. Zhang, H. Oh, R.A. Burton, J.W. Burgner, R.L. Geahlen, and C.B. Post. Tyr130 phosphorylation triggers Syk release from antigen receptor by long-distance conformational uncoupling. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(33):11760–11765.

[10] W. Wriggers, S. Chakravarty, and P.A. Jennings. Control of protein functional dynamics by peptide linkers. *Biopolymers*, 2005, 80(6):736–746.

[11] Y.E. Shapiro. NMR spectroscopy on domain dynamics in biomacromolecules. *Progress in Biophysics and Molecular Biology*, 2013, 112(3):58–117.

[12] C. Göbl, T. Madl, B. Simon, and M. Sattler. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2014, 80:26–63.

[13] P. Bernadó, K. Modig, P. Grela, D.I. Svergun, M. Tchorzewski, M. Pons, and M. Akke. Structure and dynamics of ribosomal protein L12: An ensemble model based on SAXS and NMR relaxation. *Biophysical Journal*, 2010, 98(10):2374–2382.

[14] F. Morcos, S. Chatterjee, C.L. McClendon, P.R. Brenner, R. Lopez-Rendon, J. Zintsmaster, M. Ercsey-Ravasz, C.R. Sweet, M.P. Jacobson, J.W. Peng, and J.A. Izaguirre. Modeling conformational ensembles of slow functional motions in Pin1-WW. *PLoS Computational Biology*, 2010, 6(12):e1001015.

[15] R.B. Fenwick, S. Esteban-Martin, and X. Salvatella. Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *European Biophysics Journal with Biophysics Letters*, 2011, 40(12):1339–1355.

[16] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal*, 2008, 275(1):1–21.

[17] S.E. Tsutakawa, G.L. Hura, K.A. Frankel, P.K. Cooper, and J.A. Tainer. Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. *Journal of Structural Biology*, 2007, 158(2):214–223.

[18] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society*, 1982, 104(17):4546–4559.

[19] M.V. Deshmukh, M. John, M. Coles, J. Peters, W. Baumeister, and H. Kessler. Inter-domain orientation and motions in VAT-N explored by residual dipolar couplings and ^{15}N backbone relaxation. *Magnetic Resonance in Chemistry*, 2006, 44:S89–S100.

[20] K.F. Hu, V. Galius, and K. Pervushin. Structural plasticity of peptidyl–prolyl isomerase sFkpA is a key to its chaperone function as revealed by solution NMR. *Biochemistry*, 2006, 45(39):11983–11991.

[21] N.B. Holland, Y. Nishimiya, S. Tsuda, and F.D. Sonnichsen. Two domains of RD3 antifreeze protein diffuse independently. *Biochemistry*, 2008, 47(22):5935–5941.

[22] J.D. Walsh, K. Meier, R. Ishima, and A.M. Gronenborn. NMR studies on domain diffusion and alignment in modular GB1 repeats. *Biophysical Journal*, 2010, 99(8):2636–2646.

[23] C.D. Mackereth, T. Madl, S. Bonnal, B. Simon, K. Zanier, A. Gasch, V. Rybin, J. Valcarcel, and M. Sattler. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature*, 2011, 475(7356):408–411.

[24] C. Schmitz, R. Vernon, G. Otting, D. Baker, and T. Huber. Protein structure determination from pseudocontact shifts using ROSETTA. *Journal of Molecular Biology*, 2012, 416(5):668–677.

[25] M.V. Petoukhov and D.I. Svergun. Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Current Opinion in Structural Biology*, 2007, 17(5):562–571.

[26] M. Pelikan, G.L. Hura, and M. Hammel. Structure and flexibility within proteins as identified through small angle X-ray scattering. *General Physiology and Biophysics*, 2009, 28(2):174–189.

[27] A. Grishaev, J. Wu, J. Trewhella, and A. Bax. Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *Journal of the American Chemical Society*, 2005, 127(47):16621–16628.

[28] I. Bertini, V. Calderone, M. Fragai, R. Jaiswal, C. Luchinat, M. Melikian, E. Mylonas, and D.I. Svergun. Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12. *Journal of the American Chemical Society*, 2008, 130(22):7011–7021.

[29] T. Madl, F. Gabel, and M. Sattler. NMR and small-angle scattering-based structural analysis of protein complexes in solution. *Journal of Structural Biology*, 2011, 173(3):472–482.

[30] M.B. Borgia, A. Borgia, R.B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, and J. Clarke. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature*, 2011, 474(7353):662–666.

[31] M. Saha and M.C. Morais. FOLD-EM: Automated fold recognition in medium- and low-resolution (4–15 Å) electron density maps. *Bioinformatics*, 2012, 28(24):3265–3273.

[32] A. Saupe and G. Englert. High-resolution nuclear magnetic resonance spectra of orientated molecules. *Physical Review Letters*, 1963, 11(10):462–464.

[33] G.E. Pake. Nuclear resonance absorption in hydrated crystals: Fine structure of the proton line. *Journal of Chemical Physics*, 1948, 16(4):327–336.

[34] A.G. Redfield. On the theory of relaxation processes. *IBM Journal of Research and Development*, 1957, 1(1):19–31.

[35] N. Tjandra and A. Bax. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, 1997, 278(5340):1111–1114.

[36] A. Bax and N. Tjandra. High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. *Journal of Biomolecular NMR*, 1997, 10(3):289–292.

[37] D.D. Ojennus, R.M. Mitton-Fry, and D.S. Wuttke. Induced alignment and measurement of dipolar couplings of an SH2 domain through direct binding with filamentous phage. *Journal of Biomolecular NMR*, 1999, 14(2):175–179.

[38] M. Rückert and G. Otting. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society*, 2000, 122(32):7793–7797.

[39] A.Y. Denisov, E. Kloser, D.G. Gray, and A.K. Mittermaier. Protein alignment using cellulose nanocrystals: Practical considerations and range of application. *Journal of Biomolecular NMR*, 2010, 47(3):195–204.

[40] S.M. Douglas, J.J. Chou, and W.M. Shih. DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(16):6644–6648.

[41] A. Bax, G. Kontaxis, and N. Tjandra. Dipolar couplings in macromolecular structure determination. *Methods in Enzymology*, 2001, 339:127–174.

[42] J.H. Prestegard, C.M. Bougault, and A.I. Kishore. Residual dipolar couplings in structure determination of biomolecules. *Chemical Reviews*, 2004, 104(8):3519–3540.

[43] N.R. Skrynnikov. Orienting molecular fragments and molecules with residual dipolar couplings. *Comptes Rendus Physique*, 2004, 5(3):359–375.

[44] M. Blackledge. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2005, 46(1):23–61.

[45] J.R. Tolman and K. Ruan. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chemical Reviews*, 2006, 106(5):1720–1736.

[46] K. Wüthrich, G. Wider, G. Wagner, and W. Braun. Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *Journal of Molecular Biology*, 1982, 155(3):311–319.

[47] X.C. Su and G. Otting. Paramagnetic labelling of proteins and oligonucleotides for NMR. *Journal of Biomolecular NMR*, 2010, 46(1):101–112.

[48] C.D. Schwieters, J.J. Kuszewski, and G.M. Clore. Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2006, 48(1):47–62.

[49] Y. Shen, R. Vernon, D. Baker, and A. Bax. De novo protein structure generation from incomplete chemical shift assignments. *Journal of Biomolecular NMR*, 2009, 43(2):63–78.

[50] I.R. Kleckner and M.P. Foster. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta-Proteins and Proteomics*, 2011, 1814(8):942–968.

[51] N.A. Lakomek, T. Carlomagno, S. Becker, C. Griesinger, and J. Meiler. A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *Journal of Biomolecular NMR*, 2006, 34(2):101–115.

[52] N.A. Lakomek, K.F.A. Walter, C. Fares, O.F. Lange, B.L. de Groot, H. Grubmuller, R. Bruschweiler, A. Munk, S. Becker, J. Meiler, and C. Griesinger. Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics. *Journal of Biomolecular NMR*, 2008, 41(3):139–155.

[53] W. Peti, J. Meiler, R. Bruschweiler, and C. Griesinger. Model-free analysis of protein backbone motion from residual dipolar couplings. *Journal of the American Chemical Society*, 2002, 124(20):5822–5833.

[54] T. Yuwen, C.B. Post, and N.R. Skrynnikov. Domain cooperativity in multidomain proteins: What can we learn from molecular alignment in anisotropic media? *Journal of Biomolecular NMR*, 2011, 51(1-2):131–150.

[55] J.S. Huston, D. Levinson, M. Mudgetthunter, M.S. Tai, J. Novotny, M.N. Margolies, R.J. Ridge, R.E. Bruccoleri, E. Haber, R. Crea, and H. Oppermann. Protein engineering of antibody binding sites: Recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 1988, 85(16):5879–5883.

[56] F. Huang and W.M. Nau. A conformational flexibility scale for amino acids in peptides. *Angewandte Chemie-International Edition*, 2003, 42(20):2269–2272.

[57] H.J. Feldman and C.W.V. Hogue. A fast method to sample real protein conformational space. *Proteins: Structure, Function, and Genetics*, 2000, 39(2):112–131.

[58] J.N. Onuchic, Z. Luthey-Schulten, and P.G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry*, 1997, 48:545–600.

[59] M. Zweckstetter. NMR: Prediction of molecular alignment from structure using the PALES software. *Nature Protocols*, 2008, 3(4):679–690.

[60] M. Ottiger, F. Delaglio, and A. Bax. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of Magnetic Resonance*, 1998, 131(2):373–378.

[61] V. Chevelkov, K. Faelber, A. Diehl, U. Heinemann, H. Oschkinat, and B. Reif. Detection of dynamic water molecules in a microcrystalline sample of the SH3 domain of α-spectrin by MAS solid-state NMR. *Journal of Biomolecular NMR*, 2005, 31(4):295–310.

[62] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.R. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C.H. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, M. Chiu, E.C. Garner, and Z. Obradovic. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, 2001, 19(1):26–59.

[63] C.J. Oldfield and A.K. Dunker. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry*, 2014, 83:553–584.

[64] V.N. Uversky. Flexible nets of malleable guardians: Intrinsically disordered chaperones in neurodegenerative diseases. *Chemical Reviews*, 2011, 111(2):1134–1166.

[65] V.N. Uversky, V. Dave, L.M. Iakoucheva, P. Malaney, S.J. Metallo, R.R. Pathak, and A.C. Joerger. Pathological unfoldomics of uncontrolled chaos: Intrinsically disordered proteins and human diseases. *Chemical Reviews*, 2014, 114(13):6844–6879.

[66] V.N. Uversky. Amyloidogenesis of natively unfolded proteins. *Current Alzheimer Research*, 2008, 5(3):260–287.

[67] V.N. Uversky. The most important thing is the tail: Multitudinous functionalities of intrinsically disordered protein termini. *FEBS Letters*, 2013, 587(13):1891–1901.

[68] P. Tompa. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*, 2005, 579(15):3346–3354.

[69] V.N. Uversky, C.J. Oldfield, and A.K. Dunker. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. *Journal of Molecular Recognition*, 2005, 18(5):343–384.

[70] C.A. Galea, Y. Wang, S.G. Sivakolundu, and R.W. Kriwacki. Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits. *Biochemistry*, 2008, 47(29):7598–7609.

[71] K. Gunasekaran, C.J. Tsai, S. Kumar, D. Zanuy, and R. Nussinov. Extended disordered proteins: Targeting function with less scaffold. *Trends in Biochemical Sciences*, 2003, 28(2):81–85.

[72] P. Tompa. The functional benefits of protein disorder. *Journal of Molecular Structure (Theochem)*, 2003, 666:361–371.

[73] Z.R. Liu and Y.Q. Huang. Advantages of proteins being disordered. *Protein Science*, 2014, 23(5):539–550.

[74] G.A. Papoian. Proteins with weakly funneled energy landscapes challenge the classical structure–function paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(38):14237–14238.

[75] V.N. Uversky and A.K. Dunker. Understanding protein non-folding. *Biochimica et Biophysica Acta-Proteins and Proteomics*, 2010, 1804(6):1231–1264.

[76] T. Mittag and J.D. Forman-Kay. Atomic-level characterization of disordered protein ensembles. *Current Opinion in Structural Biology*, 2007, 17(1):3–14.

[77] C.K. Fisher and C.M. Stultz. Constructing ensembles for intrinsically disordered proteins. *Current Opinion in Structural Biology*, 2011, 21(3):426–431.

[78] V.N. Uversky. Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta-Proteins and Proteomics*, 2013, 1834(5):932–951.

[79] X. Deng, J. Eickholt, and J.L. Cheng. A comprehensive overview of computational protein disorder prediction methods. *Molecular Biosystems*, 2012, 8(1):114–121.

[80] C.J. Oldfield, B. Xue, Y.Y. Van, E.L. Ulrich, J.L. Markley, A.K. Dunker, and V.N. Uversky. Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochimica et Biophysica Acta-Proteins and Proteomics*, 2013, 1834(2):487–498.

[81] J. Yan, M.J. Mizianty, P.L. Filipow, V.N. Uversky, and L. Kurgan. RAPID: Fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale. *Biochimica et Biophysica Acta-Proteins and Proteomics*, 2013, 1834(8):1671–1680.

[82] M. Sickmeier, J.A. Hamilton, T. LeGall, V. Vacic, M.S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V.N. Uversky, Z. Obradovic, and A.K. Dunker. DisProt: The database of disordered proteins. *Nucleic Acids Research*, 2007, 35:D786–D793.

[83] P. Bernadó and D.I. Svergun. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Molecular Biosystems*, 2012, 8(1):151–167.

[84] P. Bernadó, E. Mylonas, M.V. Petoukhov, M. Blackledge, and D.I. Svergun. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*, 2007, 129(17):5656–5664.

[85] P. Bernadó and M. Blackledge. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophysical Journal*, 2009, 97(10):2839–2845.

[86] M. Brucale, B. Schuler, and B. Samori. Single-molecule studies of intrinsically disordered proteins. *Chemical Reviews*, 2014, 114(6):3281–3317.

[87] C. Mendoza, F. Figueirido, and M.L. Tasayco. DSC studies of a family of natively disordered fragments from *Escherichia coli* thioredoxin: Surface burial in intrinsic coils. *Biochemistry*, 2003, 42(11):3349–3358.

[88] V. Receveur-Brechot, J.M. Bourhis, V.N. Uversky, B. Canard, and S. Longhi. Assessing protein disorder and induced folding. *Proteins: Structure, Function, and Bioinformatics*, 2006, 62(1):24–45.

[89] V.N. Uversky. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, 1993, 32(48):13288–13298.

[90] V.N. Uversky. Natively unfolded proteins: A point where biology waits for physics. *Protein Science*, 2002, 11(4):739–756.

[91] M. Kjaergaard, V. Iesmantavicius, and F.M. Poulsen. The interplay between transient α-helix formation and side chain rotamer distributions in disordered proteins probed by methyl chemical shifts. *Protein Science*, 2011, 20(12):2023–2034.

[92] B.W. Luan, N. Lyle, R.V. Pappu, and D.P. Raleigh. Denatured state ensembles with the same radii of gyration can form significantly different long-range contacts. *Biochemistry*, 2014, 53(1):39–47.

[93] N.Y. Haba, R. Gross, J. Novacek, H. Shaked, L. Zidek, M. Barda-Saad, and J.H. Chill. NMR determines transient structure and dynamics in the disordered C-terminal domain of WASp interacting protein. *Biophysical Journal*, 2013, 105(2):481–493.

[94] V. Iesmantavicius, M.R. Jensen, V. Ozenne, M. Blackledge, F.M. Poulsen, and M. Kjaergaard. Modulation of the intrinsic helix propensity of an intrinsically disordered protein reveals long-range helix–helix interactions. *Journal of the American Chemical Society*, 2013, 135(27):10155–10163.

[95] R. Konrat. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance*, 2014, 241:74–85.

[96] J. Nováček, N.Y. Haba, J.H. Chill, L. Žídek, and V. Sklenar. 4D non-uniformly sampled HCBCACON and $^{1}J(\mathrm{NC}^{\alpha})$-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *Journal of Biomolecular NMR*, 2012, 53(2):139–148.

[97] A. Zawadzka-Kazimierczuk, W. Kozminski, H. Sanderova, and L. Krasny. High dimensional and high resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins. *Journal of Biomolecular NMR*, 2012, 52(4):329–337.

[98] W. Bermel, I.C. Felli, L. Gonnelli, W. K.ozminski, A. Piai, R. Pierattelli, and A. Zawadzka-Kazimierczuk. High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *Journal of Biomolecular NMR*, 2013, 57(4):353–361.

[99] J. Nováček, L. Janda, R. Dopitova, L. Žídek, and V. Sklenar. Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: Transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. *Journal of Biomolecular NMR*, 2013, 56(4):291–301.

[100] V.Y. Orekhov and V.A. Jaravine. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2011, 59(3):271–292.

[101] J. Wen, J.H. Wu, and P. Zhou. Sparsely sampled high-resolution 4-D experiments for efficient backbone resonance assignment of disordered proteins. *Journal of Magnetic Resonance*, 2011, 209(1):94–100.

[102] L. Isaksson, M. Mayzel, M. Saline, A. Pedersen, J. Rosenlow, B. Brutscher, B.G. Karlsson, and V.Y. Orekhov. Highly efficient NMR assignment of intrinsically disordered proteins: Application to B- and T cell receptor domains. *PLoS ONE*, 2013, 8(5):e62947.

[103] X. Liu and D.W. Yang. HN(CA)N and HN(COCA)N experiments for assignment of large disordered proteins. *Journal of Biomolecular NMR*, 2013, 57(2):83–89.

[104] M. Hellman, H. Piirainen, V.P. Jaakola, and P. Permi. Bridge over troubled proline: Assignment of intrinsically disordered proteins using (HCA)CON(CAN)H and (HCA)N(CA)CO(N)H experiments concomitantly with HNCO and i(HCA)CO(CA)NH. *Journal of Biomolecular NMR*, 2014, 58(1):49–60.

[105] D. Pantoja-Uceda and J. Santoro. New ^{13}C-detected experiments for the assignment of intrinsically disordered proteins. *Journal of Biomolecular NMR*, 2014, 59(1):43–50.

[106] R.L. Narayanan, U.H.N. Durr, S. Bibow, J. Biernat, E. Mandelkow, and M. Zweckstetter. Automatic assignment of the intrinsically disordered protein Tau with 441-residues. *Journal of the American Chemical Society*, 2010, 132(34):11906–11907.

[107] J.G. Reddy and R.V. Hosur. A reduced dimensionality NMR pulse sequence and an efficient protocol for unambiguous assignment in intrinsically disordered proteins. *Journal of Biomolecular NMR*, 2014, 59(3):199–210.

[108] P. Schanda, Ē. Kupče, and B. Brutscher. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR*, 2005, 33(4):199–211.

[109] I. Bertini, I.C. Felli, L. Gonnelli, M.V.V. Kumar, and R. Pierattelli. High-resolution characterization of intrinsic disorder in proteins: Expanding the suite of ^{13}C-detected NMR spectroscopy experiments to determine key observables. *Chembiochem*, 2011, 12(15):2347–2352.

[110] W. Bermel, I. Bertini, J. Chill, I.C. Felli, N. Haba, M.V.V. Kumar, and R. Pierattelli. Exclusively heteronuclear ^{13}C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs). *Chembiochem*, 2012, 13(16):2425–2432.

[111] F.L.V. Gray, M.J. Murai, J. Grembecka, and T. Cierpicki. Detection of disordered regions in globular proteins using ^{13}C-detected NMR. *Protein Science*, 2012, 21(12):1954–1960.

[112] W. Bermel, I. Bertini, I.C. Felli, L. Gonnelli, W. Kozminski, A. Piai, R. Pierattelli, and J. Stanek. Speeding up sequence specific assignment of IDPs. *Journal of Biomolecular NMR*, 2012, 53(4):293–301.

[113] J. Nováček, L. Žídek, and V. Sklenar. Toward optimal-resolution NMR of intrinsically disordered proteins. *Journal of Magnetic Resonance*, 2014, 241:41–52.

[114] J.W. Peng. Exposing the moving parts of proteins with NMR spectroscopy. *Journal of Physical Chemistry Letters*, 2012, 3(8):1039–1051.

[115] M. Hellman, H. Tossavainen, P. Rappu, J. Heino, and P. Permi. Characterization of intrinsically disordered prostate associated gene (PAGE5) at single residue resolution by NMR spectroscopy. *PLoS ONE*, 2011, 6(11):e26633.

[116] P. Kaderavek, V. Zapletal, A. Rabatinova, L. Krasny, V. Sklenar, and L. Žídek. Spectral density mapping protocols for analysis of molecular motions in disordered proteins. *Journal of Biomolecular NMR*, 2014, 58(3):193–207.

[117] P. Bernadó, C.W. Bertoncini, C. Griesinger, M. Zweckstetter, and M. Blackledge. Defining long-range order and local disorder in native α-synuclein using residual dipolar couplings. *Journal of the American Chemical Society*, 2005, 127(51):17968–17969.

[118] V. Ozenne, R. Schneider, M.X. Yao, J.R. Huang, L. Salmon, M. Zweckstetter, M.R. Jensen, and M. Blackledge. Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *Journal of the American Chemical Society*, 2012, 134(36):15138–15148.

[119] F. Sziegat, R. Silvers, M. Hahnke, M.R. Jensen, M. Blackledge, J. Wirmer-Bartoschek, and H. Schwalbe. Disentangling the coil: Modulation of conformational and dynamic properties by site-directed mutation in the non-native state of hen egg white lysozyme. *Biochemistry*, 2012, 51(16):3361–3372.

[120] P.M. Keizers and M. Ubbink. Paramagnetic tagging for protein structure and dynamics analysis. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2011, 58(1-2):88–96.

[121] Y. Xue and N.R. Skrynnikov. Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, ^{15}N relaxation, and molecular dynamics simulations: How fast is segmental diffusion in denatured ubiquitin? *Journal of the American Chemical Society*, 2011, 133(37):14614–14628.

[122] L. Salmon, G. Nodet, V. Ozenne, G.W. Yin, M.R. Jensen, M. Zweckstetter, and M. Blackledge. NMR characterization of long-range order in intrinsically disordered proteins. *Journal of the American Chemical Society*, 2010, 132(24):8407–8418.

[123] K.P. Wu, S. Kim, D.A. Fela, and J. Baum. Characterization of conformational and dynamic properties of natively unfolded human and mouse α-synuclein ensembles by NMR: Implication for aggregation. *Journal of Molecular Biology*, 2008, 378(5):1104–1115.

[124] S. Meier, M. Blackledge, and S. Grzesiek. Conformational distributions of unfolded polypeptides from novel NMR techniques. *Journal of Chemical Physics*, 2008, 128(5).

[125] M.R. Jensen, L. Salmon, G. Nodet, and M. Blackledge. Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *Journal of the American Chemical Society*, 2010, 132(4):1270–1271.

[126] K.A. Heisel and V.V. Krishnan. NMR based solvent exchange experiments to understand the conformational preference of intrinsically disordered proteins using FG-nucleoporin peptide as a model. *Biopolymers*, 2014, 102(1):69–77.

[127] J. Roche, J.F. Ying, A.S. Maltsev, and A. Bax. Impact of hydrostatic pressure on an intrinsically disordered protein: A high-pressure NMR study of α-synuclein. *Chembiochem*, 2013, 14(14):1754–1761.

[128] C. Bracken, L.M. Iakoucheva, P.R. Rorner, and A.K. Dunker. Combining prediction, computation and experiment for the characterization of protein disorder. *Current Opinion in Structural Biology*, 2004, 14(5):570–576.

[129] G. Nodet, L. Salmon, V. Ozenne, S. Meier, M.R. Jensen, and M. Blackledge. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *Journal of the American Chemical Society*, 2009, 131(49):17908–17918.

[130] V. Ozenne, F. Bauer, L. Salmon, J.R. Huang, M.R. Jensen, S. Segard, P. Bernadó, C. Charavay, and M. Blackledge. *Flexible-meccano*: A tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, 2012, 28(11):1463–1470.

[131] R. Schneider, J.R. Huang, M.X. Yao, G. Communie, V. Ozenne, L. Mollica, L. Salmon, M.R. Jensen, and M. Blackledge. Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Molecular Biosystems*, 2012, 8(1):58–68.

[132] T. Yuwen and N.R. Skrynnikov. Proton-decoupled CPMG: A better experiment for measuring ^{15}N R_2 relaxation in disordered proteins. *Journal of Magnetic Resonance*, 2014, 241:155–169.

[133] T. Yuwen and N.R. Skrynnikov. CP-HISQC: A better version of HSQC experiment for intrinsically disordered proteins under physiological conditions. *Journal of Biomolecular NMR*, 2014, 58(3):175–192.

[134] P. Allard, M. Helgstrand, and T. Hard. The complete homogeneous master equation for a heteronuclear two-spin system in the basis of Cartesian product operators. *Journal of Magnetic Resonance*, 1998, 134(1):7–16.

[135] D.M. Korzhnev, N.R. Skrynnikov, O. Millet, D.A. Torchia, and L.E. Kay. An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *Journal of the American Chemical Society*, 2002, 124(36):10743–10753.

[136] L.E. Kay, L.K. Nicholson, F. Delaglio, A. Bax, and D.A. Torchia. Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T_1 and T_2 values in proteins. *Journal of Magnetic Resonance*, 1992, 97(2):359–375.

[137] J.A. Jones, P. Hodgkinson, A.L. Barker, and P.J. Hore. Optimal sampling strategies for the measurement of spin–spin relaxation times. *Journal of Magnetic Resonance*, 1996, 113(1):25–34.

[138] A.J. Shaka, C.J. Lee, and A. Pines. Iterative schemes for bilinear operators; application to spin decoupling. *Journal of Magnetic Resonance*, 1988, 77(2):274–293.

[139] D.M. Korzhnev, E.V. Tischenko, and A.S. Arseniev. Off-resonance effects in ^{15}N T_2 CPMG measurements. *Journal of Biomolecular NMR*, 2000, 17(3):231–237.

[140] J. Iwahara, Y.S. Jung, and G.M. Clore. Heteronuclear NMR spectroscopy for lysine NH_3 groups in proteins: Unique effect of water exchange on ^{15}N transverse relaxation. *Journal of the American Chemical Society*, 2007, 129(10):2971–2980.

[141] S. Gil, T. Hosek, Z. Solyom, R. Kummerle, B. Brutscher, R. Pierattelli, and I.C. Felli. NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. *Angewandte Chemie-International Edition*, 2013, 52(45):11808–11812.

[142] J. Janin and A.M.J.J. Bonvin. Protein–protein interactions. *Current Opinion in Structural Biology*, 2013, 23(6):859–861.

[143] K. Kahn, K.W. Plaxco, and M. Zourob. Principles of Biomolecular Recognition. *Recognition Receptors in Biosensors*, 2010, pages 3–45.

[144] M. Hoefling and K.E. Gottschalk. Barnase–barstar: From first encounter to final complex. *Journal of Structural Biology*, 2010, 171(1):52–63.

[145] M. Ubbink. Dynamics in transient complexes of redox proteins. *Biochemical Society Transactions*, 2012, 40:415–418.

[146] J. Janin. The kinetics of protein–protein recognition. *Proteins: Structure, Function, and Genetics*, 1997, 28(2):153–161.

[147] M. Ubbink. The courtship of proteins: Understanding the encounter complex. *FEBS Letters*, 2009, 583(7):1060–1066.

[148] Q. Bashir, S. Scanu, and M. Ubbink. Dynamics in electron transfer protein complexes. *FEBS Journal*, 2011, 278(9):1391–1400.

[149] J. Schilder and M. Ubbink. Formation of transient protein complexes. *Current Opinion in Structural Biology*, 2013, 23(6):911–918.

[150] J.Y. Guan, J.M. Foerster, J.W. Drijfhout, M. Timmer, A. Blok, G.M. Ullmann, and M. Ubbink. An ensemble of rapidly interconverting orientations in electrostatic protein–peptide complexes characterized by NMR spectroscopy. *Chembiochem*, 2014, 15(4):556–566.

[151] A. Spaar and V. Helms. Free energy landscape of protein–protein encounter resulting from Brownian dynamics simulations of barnase:barstar. *Journal of Chemical Theory and Computation*, 2005, 1(4):723–736.

[152] J. Schluttig, D. Alamanova, V. Helms, and U.S. Schwarz. Dynamics of protein–protein encounter: A Langevin equation approach with reaction patches. *Journal of Chemical Physics*, 2008, 129(15).

[153] G. Schreiber, G. Haran, and H.X. Zhou. Fundamental aspects of protein–protein association kinetics. *Chemical Reviews*, 2009, 109(3):839–860.

[154] R. Alsallaq and H.X. Zhou. Electrostatic rate enhancement and transient complex of protein–protein association. *Proteins: Structure, Function, and Bioinformatics*, 2008, 71(1):320–335.

[155] S. Qin, X.D. Pang, and H.X. Zhou. Automated prediction of protein association rate constants. *Structure*, 2011, 19(12):1744–1751.

[156] D. Ganguly, S. Otieno, B. Waddell, L. Iconaru, R.W. Kriwacki, and J.H. Chen. Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins. *Journal of Molecular Biology*, 2012, 422(5):674–684.

[157] M. Ubbink. Complexes of photosynthetic redox proteins studied by NMR. *Photosynthesis Research*, 2004, 81(3):277–287.

[158] A. Perrakis, A. Musacchio, S. Cusack, and C. Petosa. Investigating a macromolecular complex: The toolkit of methods. *Journal of Structural Biology*, 2011, 175(2):106–112.

[159] J. Piehler. Spectroscopic techniques for monitoring protein interactions in living cells. *Current Opinion in Structural Biology*, 2014, 24:54–62.

[160] A.S.J. Melquiond, E. Karaca, P.L. Kastritis, and A.M.J.J. Bonvin. Next challenges in protein–protein docking: From proteome to interactome and beyond. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 2012, 2(4):642–651.

[161] A. Szilagyi and Y. Zhang. Template-based structure modeling of protein–protein interactions. *Current Opinion in Structural Biology*, 2014, 24:10–23.

[162] J.R. Perkins, I. Diboun, B.H. Dessailly, J.G. Lees, and C. Orengo. Transient protein–protein interactions: Structural, functional, and network properties. *Structure*, 2010, 18(10):1233–1243.

[163] S.E.A. Ozbabacan, H.B. Engin, A. Gursoy, and O. Keskin. Transient protein–protein interactions. *Protein Engineering Design and Selection*, 2011, 24(9):635–648.

[164] M.M. Pierce, C.S. Raman, and B.T. Nall. Isothermal titration calorimetry of protein–protein interactions. *Methods*, 1999, 19(2):213–221.

[165] A.J. Rowe. Ultra-weak reversible protein–protein interactions. *Methods*, 2011, 54(1):157–166.

[166] A.M.J.J. Bonvin, R. Boelens, and R. Kaptein. NMR analysis of protein interactions. *Current Opinion in Chemical Biology*, 2005, 9(5):501–508.

[167] K. Takeuchi and G. Wagner. NMR studies of protein interactions. *Current Opinion in Structural Biology*, 2006, 16(1):109–117.

[168] L. Fielding. NMR methods for the determination of protein–ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2007, 51(4):219–242.

[169] L.Y. Lian. NMR studies of weak protein–protein interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2013, 71:59–72.

[170] M.A.S. Hass and M. Ubbink. Structure determination of protein–protein complexes with long-range anisotropic paramagnetic NMR restraints. *Current Opinion in Structural Biology*, 2014, 24:45–53.

[171] G.M. Clore, C. Tang, and J. Iwahara. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Current Opinion in Structural Biology*, 2007, 17(5):603–616.

[172] G.M. Clore. Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation. *Protein Science*, 2011, 20(2):229–246.

[173] D.M. Korzhnev and L.E. Kay. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Accounts of Chemical Research*, 2008, 41(3):442–451.

[174] D.M. Korzhnev, X. Salvatella, M. Vendruscolo, A.A. Di Nardo, A.R. Davidson, C.M. Dobson, and L.E. Kay. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature*, 2004, 430(6999):586–590.

[175] I. Bezsonova, M.C. Bruce, S. Wiesner, H. Lin, D. Rotin, and J.D. Forman-Kay. Interactions between the three CIN85 SH3 domains and ubiquitin: Implications for CIN85 ubiquitination. *Biochemistry*, 2008, 47(34):8937–8949.

[176] L. Salmon, J.L.O. Roldan, E. Lescop, A. Licinio, N. van Nuland, M.R. Jensen, and M. Blackledge. Structure, dynamics, and kinetics of weak protein–protein complexes from NMR spin relaxation measurements of titrated solutions. *Angewandte Chemie-International Edition*, 2011, 50(16):3755–3759.

[177] J. Vaynberg, T. Fukuda, K. Chen, O. Vinogradova, A. Velyvis, Y.Z. Tu, L. Ng, C.Y. Wu, and J. Qin. Structure of an ultraweak protein–protein complex and its crucial role in regulation of cell morphology and motility. *Molecular Cell*, 2005, 17(4):513–523.

[178] M.R. Jensen, J.L. Ortega-Roldan, L. Salmon, N. van Nuland, and M. Blackledge. Characterizing weak protein–protein complexes by NMR residual dipolar couplings. *European Biophysics Journal with Biophysics Letters*, 2011, 40(12):1371–1381.

[179] X.F. Xu, W.G. Reinle, F. Hannemann, P.V. Konarev, D.I. Svergun, R. Bernhardt, and M. Ubbink. Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy. *Journal of the American Chemical Society*, 2008, 130(20):6395–6403.

[180] M. Karplus and J.A. McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 2002, 9(9):646–652.

[181] J.A. McCammon, B.R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 1977, 267(5612):585–590.

[182] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y.B. Shan, and W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 2010, 330(6002):341–346.

[183] Y. Xue, J.M. Ward, T. Yuwen, I.S. Podkorytov, and N.R. Skrynnikov. Microsecond time-scale conformational exchange in proteins: Using long molecular dynamics trajectory to simulate NMR relaxation dispersion data. *Journal of the American Chemical Society*, 2012, 134(5):2555–2562.

[184] J.E. Stone, D.J. Hardy, I.S. Ufimtsev, and K. Schulten. GPU-accelerated molecular modeling coming of age. *Journal of Molecular Graphics and Modelling*, 2010, 29(2):116–125.

[185] R.O. Dror, R.M. Dirks, J.P. Grossman, H.F. Xu, D.E. Shaw, and D.C. Rees. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, 2012, 41:429–452.

[186] J.R. Allison, S. Hertig, J.H. Missimer, L.J. Smith, M.O. Steinmetz, and J. Dolenc. Probing the structure and dynamics of proteins by combining molecular dynamics simulations and experimental NMR data. *Journal of Chemical Theory and Computation*, 2012, 8(10):3430–3444.

[187] S. Esteban-Martin, R.B. Fenwick, and X. Salvatella. Synergistic use of NMR and MD simulations to study the structural heterogeneity of proteins. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 2012, 2(3):466–478.

[188] P. Robustelli, K. Kohlhoff, A. Cavalli, and M. Vendruscolo. Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure*, 2010, 18(8):923–933.

[189] A.C. Stelzer, A.T. Frank, M.H. Bailor, I. Andricioaei, and H.M. Al-Hashimi. Constructing atomic-resolution RNA structural ensembles using MD and motionally decoupled NMR RDCs. *Methods*, 2009, 49(2):167–173.

[190] J.R. Allison, P. Varnai, C.M. Dobson, and M. Vendruscolo. Determination of the free energy landscape of α-synuclein using spin label nuclear magnetic resonance measurements. *Journal of the American Chemical Society*, 2009, 131(51):18314–18326.

[191] D.W. Li and R. Bruschweiler. NMR-based protein potentials. *Angewandte Chemie-International Edition*, 2010, 49(38):6778–6780.

[192] D.W. Li and R. Bruschweiler. Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. *Journal of Chemical Theory and Computation*, 2011, 7(6):1773–1782.

[193] S.A. Showalter and R. Bruschweiler. Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field. *Journal of Chemical Theory and Computation*, 2007, 3(3):961–975.

[194] L. Wickstrom, A. Okur, and C. Simmerling. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophysical Journal*, 2009, 97(3):853–856.

[195] D.W. Li and R. Bruschweiler. Certification of molecular dynamics trajectories with NMR chemical shifts. *Journal of Physical Chemistry Letters*, 2010, 1(1):246–248.

[196] A.N. Koller, H. Schwalbe, and H. Gohlke. Starting structure dependence of NMR order parameters derived from MD simulations: Implications for judging force-field quality. *Biophysical Journal*, 2008, 95(1):L4–L6.

[197] X.D. Wu, B. Knudsen, S.M. Feller, J. Zheng, A. Sali, D. Cowburn, H. Hanafusa, and J. Kuriyan. Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure*, 1995, 3(2):215–226.

[198] M. Ahmad, W. Gu, and V. Helms. Mechanism of fast peptide recognition by SH3 domains. *Angewandte Chemie-International Edition*, 2008, 47(40):7626–7630.

[199] A.N. Volkov, Q. Bashir, J.A.R. Worrall, G.M. Ullmann, and M. Ubbink. Shifting the equilibrium between the encounter state and the specific form of a protein complex by interfacial point mutations. *Journal of the American Chemical Society*, 2010, 132(33):11487–11495.

[200] B.S. Knudsen, J. Zheng, S.M. Feller, J.P. Mayer, S.K. Burrell, D. Cowburn, and H. Hanafusa. Affinity and specificity requirements for the first Src homology 3 domain of the Crk proteins. *EMBO Journal*, 1995, 14(10):2191–2198.

[201] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 2006, 65(3):712–725.

[202] R.B. Best and G. Hummer. Optimized molecular dynamics force fields applied to the helix–coil transition of polypeptides. *Journal of Physical Chemistry B*, 2009, 113(26):9004–9015.

[203] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, and D.E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, 2010, 78(8):1950–1958.

[204] C.M. Baker, P.E.M. Lopes, X. Zhu, B. Roux, and A.D. MacKerell. Accurate calculation of hydration free energies using pair-specific Lennard–Jones parameters in the CHARMM Drude polarizable force field. *Journal of Chemical Theory and Computation*, 2010, 6(4):1181–1198.

[205] R.M. Venable, Y. Luo, K. Gawrisch, B. Roux, and R.W. Pastor. Simulations of anionic lipid membranes: Development of interaction-specific ion parameters and validation using NMR data. *Journal of Physical Chemistry B*, 2013, 117(35):10183–10192.

[206] D.E. Chapman, J.K. Steck, and P.S. Nerenberg. Optimizing protein–protein van der Waals interactions for the AMBER ff9x/ff12 force field. *Journal of Chemical Theory and Computation*, 2014, 10(1):273–281.

[207] J. Kahlen, L. Salimi, M. Sulpizi, C. Peter, and D. Donadio. Interaction of charged amino-acid side chains with ions: An optimization strategy for classical force fields. *Journal of Physical Chemistry B*, 2014, 118(14):3960–3972.

[208] Y. Xue, T. Yuwen, F. Zhu, and N.R. Skrynnikov. Role of electrostatic interactions in binding of peptides and intrinsically disordered proteins to their folded targets. 1. NMR and MD characterization of the complex between the c-Crk N-SH3 domain and the peptide Sos. *Biochemistry*, 2014, 53(41):6473–6495.

[209] J.W.H. Schymkowitz, F. Rousseau, I.C. Martins, J. Ferkinghoff-Borg, F. Stricher, and L. Serrano. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(29):10147–10152.

[210] Y. Kobashigawa, M. Sakai, M. Naito, M. Yokochi, H. Kumeta, Y. Makino, K. Ogura, S. Tanaka, and F. Inagaki. Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. *Nature Structural and Molecular Biology*, 2007, 14(6):503–510.

[211] S.M. Feller. Crk family adaptors—signalling complex formation and biological roles. *Oncogene*, 2001, 20(44):6348–6371.

[212] S. Akakura, B. Kar, S. Singh, L. Cho, N. Tibrewal, R. Sanokawa-Akakura, C. Reichman, K.S. Ravichandran, and R.B. Birge. C-terminal SH3 domain of Crkll regulates the assembly and function of the DOCK180/ELMO Rac-GEF. *Journal of Cellular Physiology*, 2005, 204(1):344–351.

[213] V. Muralidharan, K. Dutta, J. Cho, M. Vila-Perello, D.P. Raleigh, D. Cowburn, and T.W. Muir. Solution structure and folding characteristics of the C-terminal SH3 domain of c-Crk-II. *Biochemistry*, 2006, 45(29):8874–8884.

[214] A. Cipres, Y.A. Abassi, and K. Vuori. Abl functions as a negative regulator of Met-induced cell motility via phosphorylation of the adapter protein CrkII. *Cellular Signalling*, 2007, 19(8):1662–1670.

[215] M.E. Peterson and E.O. Long. Inhibitory receptor signaling via tyrosine phosphorylation of the adaptor Crk. *Immunity*, 2008, 29(4):578–588.

[216] S.H. Bae, H.J. Dyson, and P.E. Wright. Prediction of the rotational tumbling time for proteins with disordered segments. *Journal of the American Chemical Society*, 2009, 131(19):6814–6821.

[217] K. Chen and N. Tjandra. Extended model free approach to analyze correlation functions of multidomain proteins in the presence of motional coupling. *Journal of the American Chemical Society*, 2008, 130(38):12745–12751.

[218] V. Wong, D.A. Case, and A. Szabo. Influence of the coupling of interdomain and overall motions on NMR relaxation. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(27):11016–11021.

[219] J.G. de la Torre, M.L. Huertas, and B. Carrasco. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophysical Journal*, 2000, 78(2):719–730.

[220] I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani, and M.A. Zoroddu. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(18):6841–6846.

[221] I. Bertini, Y.K. Gupta, C. Luchinat, G. Parigi, M. Peana, L. Sgheri, and J. Yuan. Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. *Journal of the American Chemical Society*, 2007, 129(42):12786–12794.

[222] T. Ikegami, L. Verdier, P. Sakhaii, S. Grimme, B. Pescatore, K. Saxena, K.M. Fiebig, and C. Griesinger. Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions. *Journal of Biomolecular NMR*, 2004, 29(3):339–349.

[223] F. Rodriguez-Castaneda, P. Haberz, A. Leonov, and C. Griesinger. Paramagnetic tagging of diamagnetic proteins for solution NMR. *Magnetic Resonance in Chemistry*, 2006, 44:S10–S16.

[224] P.H.J. Keizers, J.F. Desreux, M. Overhand, and M. Ubbink. Increased paramagnetic effect of a lanthanide protein probe by two-point attachment. *Journal of the American Chemical Society*, 2007, 129(30):9292–9293.

[225] A. Musacchio, M. Noble, R. Pauptit, R. Wierenga, and M. Saraste. Crystal structure of a Src-homology 3 (SH3) domain. *Nature*, 1992, 359(6398):851–855.

[226] S.R. Hartmann and E.L. Hahn. Nuclear double resonance in the rotating frame. *Physical Review*, 1962, 128(5):2042–2053.

[227] T. Gullion. Introduction to rotational-echo, double-resonance NMR. *Concepts in Magnetic Resonance*, 1998, 10(5):277–289.

[228] J.S. Waugh, L.M. Huber, and U. Haeberlen. Approach to high-resolution NMR in solids. *Physical Review Letters*, 1968, 20(5):180–182.

[229] U. Haeberlen and J.S. Waugh. Coherent averaging effects in magnetic resonance. *Physical Review*, 1968, 175(2):453–467.

[230] E.R. Andrew. Magic angle spinning in solid state NMR spectroscopy. *Philosophical Transactions of the Royal Society A*, 1981, 299(1452):505–520.

[231] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, and J.H. Prestegard. Structural and dynamic analysis of residual dipolar coupling data for proteins. *Journal of the American Chemical Society*, 2001, 123(7):1416–1424.

[232] J.A. Losonczi, M. Andrec, M.W.F. Fischer, and J.H. Prestegard. Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance*, 1999, 138(2):334–342.

[233] G.M. Clore, A.M. Gronenborn, and A. Bax. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *Journal of Magnetic Resonance*, 1998, 133(1):216–221.

[234] J.M. Ward and N.R. Skrynnikov. Very large residual dipolar couplings from deuterated ubiquitin. *Journal of Biomolecular NMR*, 2012, 54(1):53–67.

[235] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 2009, 30(10):1545–1614.

[236] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, and G.M. Clore. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, 2003, 160(1):65–73.

[237] B.N. Dominy and C.L. Brooks. Development of a Generalized Born model parametrization for proteins and nucleic acids. *Journal of Physical Chemistry B*, 1999, 103(18):3765–3773.

[238] F.J. Blanco, A.R. Ortiz, and L. Serrano. ^{1}H and ^{15}N NMR assignment and solution structure of the SH3 domain of spectrin: Comparison of unrefined and refined structure sets with the crystal structure. *Journal of Biomolecular NMR*, 1997, 9(4):347–357.

[239] C.D. Schwieters and G.M. Clore. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *Journal of Magnetic Resonance*, 2001, 152(2):288–302.

[240] J.D. Bryngelson and P.G. Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 1987, 84(21):7524–7528.

[241] E.I. Shakhnovich and A.M. Gutin. Formation of unique structure in polypeptide chains: Theoretical investigation with the aid of a replica approach. *Biophysical Chemistry*, 1989, 34(3):187–199.

[242] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, and P.G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Genetics*, 1995, 21(3):167–195.

[243] M.P. Moreno-Murciano, D. Monleon, C. Marcinkiewcz, J.J. Calvete, and B. Celda. NMR solution structure of the non-RGD disintegrin obtustatin. *Journal of Molecular Biology*, 2003, 329(1):135–145.

[244] K. Futterer, J. Wong, R.A. Grucza, A.C. Chan, and G. Waksman. Structural basis for syk tyrosine kinase ubiquity in signal transduction pathways revealed by the crystal structure of its regulatory SH2 domains bound to a dually phosphorylated ITAM peptide. *Journal of Molecular Biology*, 1998, 281(3):523–537.

[245] K.O. Hakansson. The structure of Mg-ATPase nucleotide-binding domain at 1.6 Å resolution reveals a unique ATP-binding motif. *Acta Crystallographica Section D*, 2009, 65:1181–1186.

[246] K. Berlin, D.P. O'Leary, and D. Fushman. Improvement and analysis of computational methods for prediction of residual dipolar couplings. *Journal of Magnetic Resonance*, 2009, 201(1):25–33.

[247] E. Freyssingeas, F. Nallet, and D. Roux. Measurement of the membrane flexibility in lamellar and "sponge" phases of the $C_{12}E_5$/hexanol/water system. *Langmuir*, 1996, 12(25):6028–6035.

[248] M. Ottiger and A. Bax. Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *Journal of Biomolecular NMR*, 1998, 12(3):361–372.

[249] K. Simon, J. Xu, C. Kim, and N. Skrynnikov. Estimating the accuracy of protein structures using residual dipolar couplings. *Journal of Biomolecular NMR*, 2005, 33(2):83–93.

[250] H. Li, A.D. Robertson, and J.H. Jensen. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins: Structure, Function, and Bioinformatics*, 2005, 61(4):704–721.

[251] D.C. Bas, D.M. Rogers, and J.H. Jensen. Very fast prediction and rationalization of pK_a values for protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics*, 2008, 73(3):765–783.

[252] M. Zweckstetter, G. Hummer, and A. Bax. Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophysical Journal*, 2004, 86(6):3444–3460.

[253] B. Wu, M. Petersen, F. Girard, M. Tessari, and S.S. Wijmenga. Prediction of molecular alignment of nucleic acids in aligned media. *Journal of Biomolecular NMR*, 2006, 35(2):103–115.

[254] M. Zweckstetter. Prediction of charge-induced molecular alignment: Residual dipolar couplings at pH 3 and alignment in surfactant liquid crystalline phases. *European Biophysics Journal with Biophysics Letters*, 2006, 35(2):170–180.

[255] N.C. Fitzkee and G.D. Rose. Reassessing random-coil statistics in unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(34):12497–12502.

[256] G.A. Lazar, J.R. Desjarlais, and T.M. Handel. De novo design of the hydrophobic core of ubiquitin. *Protein Science*, 1997, 6(6):1167–1178.

[257] C.A. Bewley. Solution structure of a cyanovirin-N:Manα1-2Manα complex: Structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure*, 2001, 9(10):931–940.

[258] D.T. Braddock, J.L. Baber, D. Levens, and G.M. Clore. Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: Solution structure of a complex between hnRNP KKH3 and single-stranded DNA. *EMBO Journal*, 2002, 21(13):3476–3485.

[259] T.S. Ulmer, B.E. Ramirez, F. Delaglio, and A. Bax. Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *Journal of the American Chemical Society*, 2003, 125(30):9179–9191.

[260] D.W. Yang and K. Nagayama. A sensitivity-enhanced method for measuring heteronuclear long-range coupling constants from the displacement of signals in two 1D subspectra. *Journal of Magnetic Resonance*, 1996, 118(1):117–121.

[261] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 1995, 6(3):277–293.

[262] A. Bax. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Science*, 2003, 12(1):1–16.

[263] M.C. Vega, J.C. Martinez, and L. Serrano. Thermodynamic and structural characterization of Asn and Ala residues in the disallowed II$'$ region of the Ramachandran plot. *Protein Science*, 2000, 9(12):2322–2328.

[264] J.C. Martinez, M.T. Pisabarro, and L. Serrano. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nature Structural Biology*, 1998, 5(8):721–729.

[265] V. Chevelkov, Y. Xue, R. Linser, N.R. Skrynnikov, and B. Reif. Comparison of solid-state dipolar couplings and solution relaxation data provides insight into protein backbone dynamics. *Journal of the American Chemical Society*, 2010, 132(14):5015–5016.

[266] S. Gaemers and A. Bax. Morphology of three lyotropic liquid crystalline biological NMR media studied by translational diffusion anisotropy. *Journal of the American Chemical Society*, 2001, 123(49):12343–12352.

[267] K.B. Briggman and J.R. Tolman. De novo determination of bond orientations and order parameters from residual dipolar couplings with high accuracy. *Journal of the American Chemical Society*, 2003, 125(34):10164–10165.

[268] B.J. van Rossum, F. Castellani, J. Pauli, K. Rehbein, J. Hollander, H.J.M. de Groot, and H. Oschkinat. Assignment of amide proton signals by combined evaluation of HN, NN and HNCA MAS-NMR correlation spectra. *Journal of Biomolecular NMR*, 2003, 25(3):217–223.

[269] L. van Dam, G. Karlsson, and K. Edwards. Morphology of magnetically aligning DMPC/DHPC aggregates—perforated sheets, not disks. *Langmuir*, 2006, 22(7):3280–3285.

[270] R. Strey, R. Schomacker, D. Roux, F. Nallet, and U. Olsson. Dilute lamellar and L_3 phases in the binary water–$C_{12}E_5$ system. *Journal of the Chemical Society-Faraday Transactions*, 1990, 86(12):2253–2261.

[271] M. Jonstromer and R. Strey. Nonionic bilayers in dilute solutions: Effect of additives. *Journal of Physical Chemistry*, 1992, 96(14):5993–6000.

[272] S. Ohnishi and D. Shortle. Observation of residual dipolar couplings in short peptides. *Proteins: Structure, Function, and Genetics*, 2003, 50(4):546–551.

[273] J.Y. Suh, M.L. Cai, and G.M. Clore. Impact of phosphorylation on structure and thermodynamics of the interaction between the N-terminal domain of enzyme I and the histidine phosphocarrier protein of the bacterial phosphotransferase system. *Journal of Biological Chemistry*, 2008, 283(27):18980–18989.

[274] Y.S. Jung, M.L. Cai, and G.M. Clore. Solution structure of the $IIA^{Chitobiose}$–$IIB^{Chitobiose}$ complex of the N,N′-diacetylchitobiose branch of the *Escherichia coli* phosphotransferase system. *Journal of Biological Chemistry*, 2010, 285(6):4173–4184.

[275] D.M. Korzhnev, T.L. Religa, W. Banachewicz, A.R. Fersht, and L.E. Kay. A transient and low-populated protein-folding intermediate at atomic resolution. *Science*, 2010, 329(5997):1312–1316.

[276] S.I. Jeon, J.H. Lee, J.D. Andrade, and P.G. Degenne. Protein–surface interactions in the presence of polyethylene oxide: I. Simplified theory. *Journal of Colloid and Interface Science*, 1991, 142(1):149–158.

[277] T. McPherson, A. Kidane, I. Szleifer, and K. Park. Prevention of protein adsorption by tethered poly(ethylene oxide) layers: Experiments and single-chain mean-field analysis. *Langmuir*, 1998, 14(1):176–186.

[278] J. Zheng, L.Y. Li, S.F. Chen, and S.Y. Jiang. Molecular simulation study of water interactions with oligo (ethylene glycol)-terminated alkanethiol self-assembled monolayers. *Langmuir*, 2004, 20(20):8931–8938.

[279] J. Zheng, L.Y. Li, H.K. Tsao, Y.J. Sheng, S.F. Chen, and S.Y. Jiang. Strong repulsive forces between protein and oligo (ethylene glycol) self-assembled monolayers: A molecular simulation study. *Biophysical Journal*, 2005, 89(1):158–166.

[280] T. Dingemans, D.J. Photinos, E.T. Samulski, A.F. Terzis, and C. Wutz. Ordering of apolar and polar solutes in nematic solvents. *Journal of Chemical Physics*, 2003, 118(15):7046–7061.

[281] K.E. Wardle, E. Carlson, D. Henderson, and R.L. Rowley. Molecular dynamics simulation of the effect of ions on a liquid–liquid interface for a partially miscible mixture. *Journal of Chemical Physics*, 2004, 120(16):7681–7688.

[282] K.A. Karraker and C.J. Radke. Disjoining pressures zeta potentials and surface tensions of aqueous non-ionic surfactant/electrolyte solutions: Theory and comparison to experiment. *Advances in Colloid and Interface Science*, 2002, 96(1-3):231–264.

[283] C. Stubenrauch and R. von Klitzing. Disjoining pressure in thin liquid foam and emulsion films—new concepts and perspectives. *Journal of Physics-Condensed Matter*, 2003, 15(27):R1197–R1232.

[284] D. Balzer. Cloud point phenomena in the phase behavior of alkyl polyglucosides in water. *Langmuir*, 1993, 9(12):3375–3384.

[285] R. Schomacker and R. Strey. Effect of ionic surfactants on nonionic bilayers: Bending elasticity of weakly charged membranes. *Journal of Physical Chemistry*, 1994, 98(14):3908–3912.

[286] H. von Berlepsch and R. de Vries. Weakly charged lamellar bilayer system: Interplay between thermal undulations and electrostatic repulsion. *European Physical Journal E*, 2000, 1(2-3):141–152.

[287] D.T. Braddock, M.L. Cai, J.L. Baber, Y. Huang, and G.M. Clore. Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. *Journal of the American Chemical Society*, 2001, 123(35):8634–8635.

[288] T.S. Ulmer, J.M. Werner, and I.D. Campbell. SH3–SH2 domain orientation in Src kinases: NMR studies of Fyn. *Structure*, 2002, 10(7):901–911.

[289] D.K.Y. Poon, S.G. Withers, and L.P. McIntosh. Direct demonstration of the flexibility of the glycosylated proline–threonine linker in the Cellulomonas fimi xylanase Cex through NMR spectroscopic analysis. *Journal of Biological Chemistry*, 2007, 282(3):2091–2100.

[290] W. Wang, J. Weng, X. Zhang, M. Liu, and M. Zhang. Creating conformational entropy by increasing interdomain mobility in ligand binding regulation: A revisit to N-terminal tandem PDZ domains of PSD-95. *Journal of the American Chemical Society*, 2009, 131(2):787–796.

[291] A.T. Alexandrescu and R.A. Kammerer. Structure and disorder in the ribonuclease S-peptide probed by NMR residual dipolar couplings. *Protein Science*, 2003, 12(10):2132–2140.

[292] P. Bernadó, L. Blanchard, P. Timmins, D. Marion, R.W.H. Ruigrok, and M. Blackledge. A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(47):17002–17007.

[293] M.R. Jensen, K. Houben, E. Lescop, L. Blanchard, R.W.H. Ruigrok, and M. Blackledge. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: Application to the molecular recognition element of Sendai virus nucleoprotein. *Journal of the American Chemical Society*, 2008, 130(25):8055–8061.

[294] C.M. Slupsky, C.M. Kay, F.C. Reinach, L.B. Smillie, and B.D. Sykes. Calcium-induced dimerization of troponin C: Mode of interaction and use of trifluoroethanol as a denaturant of quaternary structure. *Biochemistry*, 1995, 34(22):7365–7375.

[295] M.E. Hodsdon and D.P. Cistola. Ligand binding alters the backbone mobility of intestinal fatty acid-binding protein as monitored by ^{15}N NMR relaxation and ^{1}H exchange. *Biochemistry*, 1997, 36(8):2278–2290.

[296] L. Zídek, M.V. Novotny, and M.J. Stone. Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nature Structural Biology*, 1999, 6(12):1118–1121.

[297] M.J. Seewald, K. Pichumani, C. Stowell, B.V. Tibbals, L. Regan, and M.J. Stone. The role of backbone conformational heat capacity in protein stability: Temperature dependent dynamics of the B1 domain of Streptococcal protein G. *Protein Science*, 2000, 9(6):1177–1193.

[298] J.M. Vinther, S.M. Kristensen, and J.J. Led. Enhanced stability of a protein with increasing temperature. *Journal of the American Chemical Society*, 2011, 133(2):271–278.

[299] L. Ragona, M. Catalano, M. Luppi, D. Cicero, T. Eliseo, J. Foote, F. Fogolari, L. Zetta, and H. Molinari. NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein. *Journal of Biological Chemistry*, 2006, 281(14):9697–9709.

[300] R. Das, M.T. Mazhab-Jafari, S. Chowdhury, S. SilDas, R. Selvaratnam, and G. Melacini. Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP. *Journal of Biological Chemistry*, 2008, 283(28):19691–19703.

[301] L.S. Yao, B. Vogeli, J.F. Ying, and A. Bax. NMR determination of amide N–H equilibrium bond length from concerted dipolar coupling measurements. *Journal of the American Chemical Society*, 2008, 130(49):16518–16519.

[302] L.S. Yao, A. Grishaev, G. Cornilescu, and A. Bax. Site-specific backbone amide ^{15}N chemical shift anisotropy tensors in a small protein from liquid crystal and cross-correlated relaxation measurements. *Journal of the American Chemical Society*, 2010, 132(12):4295–4309.

[303] D.F. Hansen, H.Q. Feng, Z. Zhou, Y.W. Bai, and L.E. Kay. Selective characterization of microsecond motions in proteins by NMR relaxation. *Journal of the American Chemical Society*, 2009, 131(44):16257–16265.

[304] F. Ferrage, D. Cowburn, and R. Ghose. Accurate sampling of high-frequency motions in proteins by steady-state ^{15}N–$\{^1$H$\}$ nuclear Overhauser effect measurements in the presence of cross-correlated relaxation. *Journal of the American Chemical Society*, 2009, 131(17):6048–6049.

[305] S. Jurt and O. Zerbe. A study on the influence of fast amide exchange on the accuracy of ^{15}N relaxation rate constants. *Journal of Biomolecular NMR*, 2012, 54(4):389–400.

[306] A.M. Mandel, M. Akke, and A.G. Palmer. Backbone dynamics of *Escherichia coli* ribonuclease HI: Correlations with structure and function in an active enzyme. *Journal of Molecular Biology*, 1995, 246(1):144–163.

[307] D.Q. Jin, M. Andrec, G.T. Montelione, and R.M. Levy. Propagation of experimental uncertainties using the Lipari–Szabo model-free analysis of protein dynamics. *Journal of Biomolecular NMR*, 1998, 12(4):471–492.

[308] N. Tjandra, S.E. Feller, R.W. Pastor, and A. Bax. Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *Journal of the American Chemical Society*, 1995, 117(50):12562–12566.

[309] L.K. Lee, M. Rance, W.J. Chazin, and A.G. Palmer. Rotational diffusion anisotropy of proteins from simultaneous analysis of ^{15}N and $^{13}C^{\alpha}$ nuclear spin relaxation. *Journal of Biomolecular NMR*, 1997, 9(3):287–298.

[310] A.L. Lee and A.J. Wand. Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation. *Journal of Biomolecular NMR*, 1999, 13(2):101–112.

[311] S.L. Chang and N. Tjandra. Temperature dependence of protein backbone motion from carbonyl ^{13}C and amide ^{15}N NMR relaxation. *Journal of Magnetic Resonance*, 2005, 174(1):43–53.

[312] S.W. Homans. Probing the binding entropy of ligand–protein interactions by NMR. *Chembiochem*, 2005, 6(9):1585–1586.

[313] S.R. Tzeng and C.G. Kalodimos. Protein activity regulation by conformational entropy. *Nature*, 2012, 488(7410):236–240.

[314] P.A. Keifer. 90° pulse width calibrations: How to read a pulse width array. *Concepts in Magnetic Resonance*, 1999, 11(3):165–180.

[315] D. Ban, A.D. Gossert, K. Giller, S. Becker, C. Griesinger, and D. Lee. Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. *Journal of Magnetic Resonance*, 2012, 221:1–4.

[316] F. Massi, E. Johnson, C.Y. Wang, M. Rance, and A.G. Palmer. NMR $R_{1\rho}$ rotating-frame relaxation with weak radio frequency fields. *Journal of the American Chemical Society*, 2004, 126(7):2247–2256.

[317] A.G. Palmer, N.J. Skelton, W.J. Chazin, P.E. Wright, and M. Rance. Suppression of the effects of cross-correlation between dipolar and anisotropic chemical shift relaxation mechanisms in the measurement of spin–spin relaxation rates. *Molecular Physics*, 1992, 75(3):699–711.

[318] S. Zinn-Justin, P. Berthault, M. Guenneugues, and H. Desvaux. Off-resonance rf fields in heteronuclear NMR: Application to the study of slow motions. *Journal of Biomolecular NMR*, 1997, 10(4):363–372.

[319] W. Myint, Q.G. Gong, and R. Ishima. Practical aspects of ^{15}N CPMG transverse relaxation experiments for proteins in solution. *Concepts in Magnetic Resonance Part A*, 2009, 34A(2):63–75.

[320] J. Pauli, B. van Rossum, H. Forster, H.J.M. de Groot, and H. Oschkinat. Sample optimization and identification of signal patterns of amino acid side chains in 2D RFDR spectra of the α-spectrin SH3 domain. *Journal of Magnetic Resonance*, 2000, 143(2):411–416.

[321] B.A. Messerle, G. Wider, G. Otting, C. Weber, and K. Wüthrich. Solvent suppression using a spin lock in 2D and 3D NMR spectroscopy with H_2O solutions. *Journal of Magnetic Resonance*, 1989, 85(3):608–613.

[322] W. Peti, L.J. Smith, C. Redfield, and H. Schwalbe. Chemical shifts in denatured proteins: Resonance assignments for denatured ubiquitin and comparisons with other denatured proteins. *Journal of Biomolecular NMR*, 2001, 19(2):153–165.

[323] A.C. Wang and A. Bax. Reparametrization of the Karplus relation for $^3J(\mathrm{H}^{\alpha}\text{–N})$ and $^3J(\mathrm{H}^{\mathrm{N}}\text{–C}')$ in peptides from uniformly $^{13}\mathrm{C}/^{15}\mathrm{N}$-enriched human ubiquitin. *Journal of the American Chemical Society*, 1995, 117(6):1810–1813.

[324] J.M. Schmidt, Y. Hua, and F. Lohr. Correlation of ^2J couplings with protein secondary structure. *Proteins: Structure, Function, and Bioinformatics*, 2010, 78(6):1544–1562.

[325] M. Ottiger, N. Tjandra, and A. Bax. Magnetic field dependent amide $^{15}\mathrm{N}$ chemical shifts in a protein–DNA complex resulting from magnetic ordering in solution. *Journal of the American Chemical Society*, 1997, 119(41):9825–9830.

[326] N.A. Farrow, O.W. Zhang, J.D. Forman-Kay, and L.E. Kay. Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry*, 1997, 36(9):2390–2402.

[327] N.R. Skrynnikov and R.R. Ernst. Detection of intermolecular chemical exchange through decorrelation of two-spin order. *Journal of Magnetic Resonance*, 1999, 137(1):276–280.

[328] G.N.B Yip and E.R.P. Zuiderweg. A phase cycle scheme that significantly suppresses offset-dependent artifacts in the R_2-CPMG $^{15}\mathrm{N}$ relaxation experiment. *Journal of Magnetic Resonance*, 2004, 171(1):25–36.

[329] D. Long, M.L. Liu, and D.W. Yang. Accurately probing slow motions on millisecond timescales with a robust NMR relaxation experiment. *Journal of the American Chemical Society*, 2008, 130(51):17629–17629.

[330] A.D. Bain, C.K. Anand, and Z.H. Nie. Exact solution of the CPMG pulse sequence with phase variation down the echo train: Application to R_2 measurements. *Journal of Magnetic Resonance*, 2011, 209(2):183–194.

[331] W. Myint, Y.F. Cai, C.A. Schiffer, and R. Ishima. Quantitative comparison of errors in $^{15}\mathrm{N}$ transverse relaxation rates measured using various CPMG phasing schemes. *Journal of Biomolecular NMR*, 2012, 53(1):13–23.

[332] Ē. Kupče, J. Boyd, and I.D. Campbell. Short selective pulses for biochemical applications. *Journal of Magnetic Resonance*, 1995, 106(3):300–303.

[333] H. Geen and R. Freeman. Band-selective radiofrequency pulses. *Journal of Magnetic Resonance*, 1991, 93(1):93–141.

[334] A. Bax, M. Ikura, L.E. Kay, D.A. Torchia, and R. Tschudin. Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. *Journal of Magnetic Resonance*, 1990, 86(2):304–318.

[335] T.J. Norwood, J. Boyd, J.E. Heritage, N. Soffe, and I.D. Campbell. Comparison of techniques for ^{1}H-detected heteronuclear ^{1}H–^{15}N spectroscopy. *Journal of Magnetic Resonance*, 1990, 87(3):488–501.

[336] A.G. Palmer, J. Cavanagh, P.E. Wright, and M. Rance. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance*, 1991, 93(1):151–170.

[337] S. Grzesiek and A. Bax. The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *Journal of the American Chemical Society*, 1993, 115(26):12593–12594.

[338] N.A. Farrow, R. Muhandiram, A.U. Singer, S.M. Pascal, C.M. Kay, G. Gish, S.E. Shoelson, T. Pawson, J.D. Forman-Kay, and L.E. Kay. Backbone dynamics of a free and a phosphopeptide-complexed Src homology-2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, 1994, 33(19):5984–6003.

[339] K. Chen and N. Tjandra. Water proton spin saturation affects measured protein backbone ^{15}N spin relaxation rates. *Journal of Magnetic Resonance*, 2011, 213(1):151–157.

[340] N.A. Lakomek, J.F. Ying, and A. Bax. Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *Journal of Biomolecular NMR*, 2012, 53(3):209–221.

[341] F. Löhr and H. Rüterjans. Detection of nitrogen–nitrogen J-couplings in proteins. *Journal of Magnetic Resonance*, 1998, 132(1):130–137.

[342] R. Ishima and D.A. Torchia. Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. *Journal of Biomolecular NMR*, 2003, 25(3):243–248.

[343] A.L. Hansen, P. Lundstrom, A. Velyvis, and L.E. Kay. Quantifying millisecond exchange dynamics in proteins by CPMG relaxation dispersion NMR using side-chain ^{1}H probes. *Journal of the American Chemical Society*, 2012, 134(6):3178–3189.

[344] G.S. Rule and T.K. Hitchens. *Fundamentals of Protein NMR Spectroscopy*. Springer, Dordrecht, 2006.

[345] M. Kjaergaard, S. Brander, and F.M. Poulsen. Random coil chemical shift for intrinsically disordered proteins: Effects of temperature and pH. *Journal of Biomolecular NMR*, 2011, 49(2):139–149.

[346] B.S. Blaum, J.A. Deakin, C.M. Johansson, A.P. Herbert, P.N. Barlow, M. Lyon, and D. Uhrin. Lysine and arginine side chains in glycosaminoglycan–protein complexes investigated by NMR, cross-linking, and mass spectrometry: A case study of the factor H–heparin interaction. *Journal of the American Chemical Society*, 2010, 132(18):6374–6381.

[347] K. Hayamizu and W.S. Price. A new type of sample tube for reducing convection effects in PGSE-NMR measurements of self-diffusion coefficients of liquid samples. *Journal of Magnetic Resonance*, 2004, 167(2):328–333.

[348] J.C. Hindman, A. Svirmickas, and M. Wood. Relaxation processes in water. A study of proton spin–lattice relaxation time. *Journal of Chemical Physics*, 1973, 59(3):1517–1522.

[349] A. Abragam. *Principles of Nuclear Magnetism*. Oxford University Press, Oxford, 1983.

[350] Y.W. Bai, J.S. Milne, L. Mayne, and S.W. Englander. Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Genetics*, 1993, 17(1):75–86.

[351] R. Orij, J. Postmus, A. Ter Beek, S. Brul, and G.J. Smits. *In vivo* measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in *Saccharomyces cerevisiae* reveals a relation between intracellular pH and growth. *Microbiology*, 2009, 155:268–278.

[352] R.L. Croke, S.M. Patil, J. Quevreaux, D.A. Kendall, and A.T. Alexandrescu. NMR determination of pK_a values in α-synuclein. *Protein Science*, 2011, 20(2):256–269.

[353] E.A. Cino, M. Karttunen, and W.Y. Choy. Effects of molecular crowding on the dynamics of intrinsically disordered proteins. *PLoS ONE*, 2012, 7(11):e49876.

[354] M. Candotti, S. Esteban-Martin, X. Salvatella, and M. Orozco. Toward an atomistic description of the urea-denatured state of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(15):5933–5938.

[355] M.K. Cho, G. Nodet, H.Y. Kim, M.R. Jensen, P. Bernadó, C.O. Fernandez, S. Becker, M. Blackledge, and M. Zweckstetter. Structural characterization of α-synuclein in an aggregation prone state. *Protein Science*, 2009, 18(9):1840–1846.

[356] R. Silvers, F. Sziegat, H. Tachibana, S. Segawa, S. Whittaker, U.L. Gunther, F. Gabel, J.R. Huang, M. Blackledge, J. Wirmer-Bartoschek, and H. Schwalbe. Modulation of structure and dynamics by disulfide bond formation in unfolded states. *Journal of the American Chemical Society*, 2012, 134(15):6846–6854.

[357] G.A. Mueller, A.M. Smith, M.D. Chapman, G.S. Rule, and D.C. Benjamin. Hydrogen exchange nuclear magnetic resonance spectroscopy mapping of antibody epitopes on the house dust mite allergen Der p 2. *Journal of Biological Chemistry*, 2001, 276(12):9359–9365.

[358] C.A. Amezcua, S.M. Harper, J. Rutter, and K.H. Gardner. Structure and interactions of PAS kinase N-terminal PAS domain: Model for intramolecular kinase regulation. *Structure*, 2002, 10(10):1349–1361.

[359] K.C. Haddad, J.L. Sudmeier, D.A. Bachovchin, and W.W. Bachovchin. α-lytic protease can exist in two separately stable conformations with different His^{57} mobilities and catalytic activities. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(4):1006–1011.

[360] C. Fernandez, C. Hilty, G. Wider, P. Guntert, and K. Wüthrich. NMR structure of the integral membrane protein OmpX. *Journal of Molecular Biology*, 2004, 336(5):1211–1221.

[361] R.L. Croke, C.O. Sallum, E. Watson, E.D. Watt, and A.T. Alexandrescu. Hydrogen exchange of monomeric α-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in *Escherichia coli*. *Protein Science*, 2008, 17(8):1434–1445.

[362] S.T.D. Hsu, C.W. Bertoncini, and C.M. Dobson. Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. *Journal of the American Chemical Society*, 2009, 131(21):7222–7223.

[363] S.N. Loh, M.S. Kay, and R.L. Baldwin. Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 1995, 92(12):5446–5450.

[364] W. Bermel, I. Bertini, I.C. Felli, M. Piccioli, and R. Pierattelli. ^{13}C-detected protonless NMR spectroscopy of proteins in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2006, 48(1):25–45.

[365] G. Pasat, J.S. Zintsmaster, and J.W. Peng. Direct ^{13}C-detection for carbonyl relaxation studies of protein dynamics. *Journal of Magnetic Resonance*, 2008, 193(2):226–232.

[366] I.C. Felli and B. Brutscher. Recent advances in solution NMR: Fast methods and heteronuclear direct detection. *Chemphyschem*, 2009, 10(9-10):1356–1368.

[367] N.D. Werbeck, J. Kirkpatrick, and D.F. Hansen. Probing arginine side-chains and their dynamics with carbon-detected NMR spectroscopy: Application to the 42 kDa human histone deacetylase 8 at high pH. *Angewandte Chemie-International Edition*, 2013, 52(11):3145–3147.

[368] N.A. Farrow, O.W. Zhang, J.D. Forman-Kay, and L.E. Kay. A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *Journal of Biomolecular NMR*, 1994, 4(5):727–734.

[369] S. Mori, C. Abeygunawardana, M.O. Johnson, and P.C.M. Vanzijl. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *Journal of Magnetic Resonance*, 1995, 108(1):94–98.

[370] P. Schanda and B. Brutscher. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, 2005, 127(22):8014–8015.

[371] S.G. Yao, M.G. Hinds, J.M. Murphy, and R.S. Norton. Exchange enhanced sensitivity gain for solvent-exchangeable protons in 2D ^{1}H–^{15}N heteronuclear correlation spectra acquired with band-selective pulses. *Journal of Magnetic Resonance*, 2011, 211(2):243–247.

[372] N.A. Farrow, O.W. Zhang, J.D. Forman-Kay, and L.E. Kay. Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry*, 1995, 34(3):868–878.

[373] V. Vidovic, L. Prongidi-Fix, B. Bechinger, and S. Werten. Production and isotope labeling of antimicrobial peptides in *Escherichia coli* by means of a novel fusion partner that enables high-yield insoluble expression and fast purification. *Journal of Peptide Science*, 2009, 15(4):278–284.

[374] O.W. Zhang and J.D. Forman-Kay. Structural characterization of folded and unfolded states of an SH3 domain equilibrium in aqueous buffer. *Biochemistry*, 1995, 34(20):6784–6794.

[375] T. Segawa, F. Kateb, L. Duma, G. Bodenhausen, and P. Pelupessy. Exchange rate constants of invisible protons in proteins determined by NMR spectroscopy. *Chembiochem*, 2008, 9(4):537–542.

[376] V. Sklenar, M. Piotto, R. Leppik, and V. Saudek. Gradient-tailored water suppression for ^{1}H–^{15}N HSQC experiments optimized to retain full sensitivity. *Journal of Magnetic Resonance*, 1993, 102(2):241–245.

[377] J.A. Marsh and J.D. Forman-Kay. Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *Journal of Molecular Biology*, 2009, 391(2):359–374.

[378] M. Bista, S.M. Freund, and A.R. Fersht. Domain-domain interactions in full-length p53 and a specific DNA complex probed by methyl NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(39):15752–15756.

[379] G.P. Connelly, Y.W. Bai, M.F. Jeng, and S.W. Englander. Isotope effects in peptide group hydrogen exchange. *Proteins: Structure, Function, and Genetics*, 1993, 17(1):87–92.

[380] O.W. Zhang, L.E. Kay, J.P. Olivier, and J.D. Forman-Kay. Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *Journal of Biomolecular NMR*, 1994, 4(6):845–858.

[381] M.H. Levitt. Heteronuclear cross polarization in liquid-state nuclear magnetic resonance mismatch compensation and relaxation behavior. *Journal of Chemical Physics*, 1991, 94(1):30–38.

[382] S. Hiller, G. Wider, T. Etezady-Esfarjani, R. Horst, and K. Wüthrich. Managing the solvent water polarization to obtain improved NMR spectra of large molecular structures. *Journal of Biomolecular NMR*, 2005, 32(1):61–70.

[383] W. Bermel, I. Bertini, I.C. Felli, and R. Pierattelli. Speeding up ^{13}C direct detection biomolecular NMR spectroscopy. *Journal of the American Chemical Society*, 2009, 131(42):15339–15345.

[384] J. Nováček, A. Zawadzka-Kazimierczuk, V. Papouskova, L. Žídek, H. Sanderova, L. Krasny, W. Kozminski, and V. Sklenar. 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *Journal of Biomolecular NMR*, 2011, 50(1):1–11.

[385] Y. Takayama, C.A. Castaneda, M. Chimenti, B. Garcia-Moreno, and J. Iwahara. Direct evidence for deprotonation of a lysine side chain buried in the hydrophobic core of a protein. *Journal of the American Chemical Society*, 2008, 130(21):6714–6715.

[386] N. Trbovic, J.H. Cho, R. Abel, R.A. Friesner, M. Rance, and A.G. Palmer. Protein side-chain dynamics and residual conformational entropy. *Journal of the American Chemical Society*, 2009, 131(2):615–622.

[387] C. Diehl, O. Engstrom, T. Delaine, M. Hakansson, S. Genheden, K. Modig, H. Leffler, U. Ryde, U.J. Nilsson, and M. Akke. Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *Journal of the American Chemical Society*, 2010, 132(41):14577–14589.

[388] A. Esadze, D.W. Li, T.Z. Wang, R. Bruschweiler, and J. Iwahara. Dynamics of lysine side-chain amino groups in a protein studied by heteronuclear ^{1}H–^{15}N NMR spectroscopy. *Journal of the American Chemical Society*, 2011, 133(4):909–919.

[389] F.H. Hu, K. Schmidt-Rohr, and M. Hong. NMR detection of pH-dependent histidine-water proton exchange reveals the conduction mechanism of a transmembrane proton channel. *Journal of the American Chemical Society*, 2012, 134(8):3703–3713.

[390] V.V. Krishnan and M. Rance. Influence of chemical exchange among homonuclear spins in heteronuclear coherence-transfer experiments in liquids. *Journal of Magnetic Resonance*, 1995, 116(1):97–106.

[391] A. Majumdar and E.R.P. Zuiderweg. Efficiencies of double-resonance and triple-resonance J cross-polarization in multidimensional NMR. *Journal of Magnetic Resonance*, 1995, 113(1):19–31.

[392] K. Zangger and I.M. Armitage. Sensitivity-enhanced detection of fast exchanging protons by an exchange-edited gradient HEHAHA-HSQC experiment. *Journal of Magnetic Resonance*, 1998, 135(1):70–75.

[393] V.N. Uversky, J.R. Gillespie, and A.L. Fink. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Genetics*, 2000, 41(3):415–427.

[394] P. Romero, Z. Obradovic, X.H. Li, E.C. Garner, C.J. Brown, and A.K. Dunker. Sequence complexity of disordered protein. *Proteins: Structure, Function, and Genetics*, 2001, 42(1):38–48.

[395] C.J. Oldfield, Y. Cheng, M.S. Cortese, C.J. Brown, V.N. Uversky, and A.K. Dunker. Comparing and combining predictors of mostly disordered proteins. *Biochemistry*, 2005, 44(6):1989–2000.

[396] V.N. Uversky and A.K. Dunker. Controlled chaos. *Science*, 2008, 322(5906):1340–1341.

[397] J.J. Ward, J.S. Sodhi, L.J. McGuffin, B.F. Buxton, and D.T. Jones. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology*, 2004, 337(3):635–645.

[398] L.M. Iakoucheva, C.J. Brown, J.D. Lawson, Z. Obradovic, and A.K. Dunker. Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology*, 2002, 323(3):573–584.

[399] H.B. Xie, S. Vucetic, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, V.N. Uversky, and Z. Obradovic. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *Journal of Proteome Research*, 2007, 6(5):1882–1898.

[400] S.A. Dames, M. Martinez-Yamout, R.N. De Guzman, H.J. Dyson, and P.E. Wright. Structural basis for Hif-1α/CBP recognition in the cellular hypoxic response. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(8):5271–5276.

[401] B. Dancheck, A.C. Naim, and W. Peti. Detailed structural characterization of unbound protein phosphatase 1 inhibitors. *Biochemistry*, 2008, 47(47):12346–12356.

[402] M.J. Ragusa, B. Dancheck, D.A. Critton, A.C. Nairn, R. Page, and W. Peti. Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nature Structural and Molecular Biology*, 2010, 17(4):459–464.

[403] L.M. Iakoucheva, P. Radivojac, C.J. Brown, T.R. O'Connor, J.G. Sikes, Z. Obradovic, and A.K. Dunker. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, 2004, 32(3):1037–1049.

[404] M. Borg, T. Mittag, T. Pawson, M. Tyers, J.D. Forman-Kay, and H.S. Chan. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(23):9650–9655.

[405] J.A. Marsh and J.D. Forman-Kay. Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical Journal*, 2010, 98(10):2383–2390.

[406] H.X. Zhou and P.A. Bates. Modeling protein association mechanisms and kinetics. *Current Opinion in Structural Biology*, 2013, 23(6):887–893.

[407] P.E. Wright and H.J. Dyson. Linking folding and binding. *Current Opinion in Structural Biology*, 2009, 19(1):31–38.

[408] A.M. Petros, D.G. Nettesheim, Y. Wang, E.T. Olejniczak, R.P. Meadows, J. Mack, K. Swift, E.D. Matayoshi, H.C. Zhang, C.B. Thompson, and S.W. Fesik. Rationale for Bcl-x_L/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Science*, 2000, 9(12):2528–2534.

[409] V. Iesmantavicius, J. Dogan, P. Jemth, K. Teilum, and M. Kjaergaard. Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angewandte Chemie-International Edition*, 2014, 53(6):1548–1551.

[410] E.R. Lacy, I. Filippov, W.S. Lewis, S. Otieno, L.M. Xiao, S. Weiss, L. Hengst, and R.W. Kriwacki. p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nature Structural and Molecular Biology*, 2004, 11(4):358–364.

[411] H.X. Zhou, X.D. Pang, and C. Lu. Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Physical Chemistry Chemical Physics*, 2012, 14(30):10466–10476.

[412] E.T.C. Wong, D. Na, and J. Gsponer. On the importance of polar interactions for complexes containing intrinsically disordered proteins. *PLoS Computational Biology*, 2013, 9(8):e1003192.

[413] D. Ganguly, W.H. Zhang, and J.H. Chen. Electrostatically accelerated encounter and folding for facile recognition of intrinsically disordered proteins. *PLoS Computational Biology*, 2013, 9(11):e1003363.

[414] T. Giorgino, I. Buch, and G. De Fabritiis. Visualizing the induced binding of SH2-phosphopeptide. *Journal of Chemical Theory and Computation*, 2012, 8(4):1171–1175.

[415] O. Dagliyan, E.A. Proctor, K.M. D'Auria, F. Ding, and N.V. Dokholyan. Structural and dynamic determinants of protein–peptide recognition. *Structure*, 2011, 19(12):1837–1845.

[416] I. Staneva, Y.Q. Huang, Z.R. Liu, and S. Wallin. Binding of two intrinsically disordered peptides to a multi-specific protein: A combined Monte Carlo and molecular dynamics study. *PLoS Computational Biology*, 2012, 8(9):e1002682.

[417] C.T. Reichman, B.J. Mayer, S. Keshav, and H. Hanafusa. The product of the cellular Crk gene consists primarily of SH2 and SH3 regions. *Cell Growth and Differentiation*, 1992, 3(7):451–460.

[418] J.A. Camarero, D. Fushman, S. Sato, I. Giriat, D. Cowburn, D.P. Raleigh, and T.W. Muir. Rescuing a destabilized protein fold through backbone cyclization. *Journal of Molecular Biology*, 2001, 308(5):1045–1062.

[419] P. Sarkar, C. Reichman, T. Saleh, R.B. Birge, and C.G. Kalodimos. Proline *cis–trans* isomerization controls autoinhibition of a signaling protein. *Molecular Cell*, 2007, 25(3):413–426.

[420] R.B. Birge, C. Kalodimos, F. Inagaki, and S. Tanaka. Crk and CrkL adaptor proteins: Networks for physiological and pathological signaling. *Cell Communication and Signaling*, 2009, 7.

[421] M. Matsuda, Y. Hashimoto, K. Muroya, H. Hasegawa, T. Kurata, S. Tanaka, S. Nakamura, and S. Hattori. CRK protein binds to two guanine nucleotide-releasing proteins for the Ras family and modulates nerve growth factor-induced activation of Ras in PC12 cells. *Molecular and Cellular Biology*, 1994, 14(8):5495–5500.

[422] S.M. Feller, B. Knudsen, and H. Hanafusa. Cellular proteins binding to the first Src homology-3 (SH3) domain of the protooncogene product c-Crk indicate Crk-specific signaling pathways. *Oncogene*, 1995, 10(8):1465–1473.

[423] C.T. Andrews and A.H. Elcock. Molecular dynamics simulations of highly crowded amino acid solutions: Comparisons of eight different force field combinations with experiment and with each other. *Journal of Chemical Theory and Computation*, 2013, 9(10):4585–4602.

[424] K.T. Debiec, A.M. Gronenborn, and L.T. Chong. Evaluating the strength of salt bridges: A comparison of current biomolecular force fields. *Journal of Physical Chemistry B*, 2014, 118(24):6561–6569.

[425] D. Tolkatchev, P. Xu, and F. Ni. Probing the kinetic landscape of transient peptide–protein interactions by use of peptide ^{15}N NMR relaxation dispersion spectroscopy: Binding of an antithrombin peptide to human prothrombin. *Journal of the American Chemical Society*, 2003, 125(41):12432–12442.

[426] K. Sugase, H.J. Dyson, and P.E. Wright. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 2007, 447(7147):1021–1025.

[427] N.L. Fawzi, J.F. Ying, R. Ghirlando, D.A. Torchia, and G.M. Clore. Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature*, 2011, 480(7376):268–272.

[428] P. Vallurupalli, G. Bouvignies, and L.E. Kay. Studying "invisible" excited protein states in slow exchange with a major state conformation. *Journal of the American Chemical Society*, 2012, 134(19):8148–8161.

[429] K.P. Wu and J. Baum. Detection of transient interchain interactions in the intrinsically disordered protein α-synuclein by NMR paramagnetic relaxation enhancement. *Journal of the American Chemical Society*, 2010, 132(16):5546–5547.

[430] C. Cheadle, Y. Ivashchenko, V. South, G.H. Searfoss, S. French, R. Howk, G.A. Ricca, and M. Jaye. Identification of a Src SH3 domain binding motif by screening a random phage display library. *Journal of Biological Chemistry*, 1994, 269(39):24034–24039.

[431] R.J. Rickles, M.C. Botfield, Z.G. Weng, J.A. Taylor, O.M. Green, J.S. Brugge, and M.J. Zoller. Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. *EMBO Journal*, 1994, 13(23):5598–5604.

[432] A.N. Volkov, M. Ubbink, and N.A.J. van Nuland. Mapping the encounter state of a transient protein complex by PRE NMR spectroscopy. *Journal of Biomolecular NMR*, 2010, 48(4):225–236.

[433] G. Schreiber and A.R. Fersht. Rapid, electrostatically assisted association of proteins. *Nature Structural Biology*, 1996, 3(5):427–431.

[434] L.N. Wafer, W.W. Streicher, S.A. McCallum, and G.I. Makhatadze. Thermodynamic and kinetic analysis of peptides derived from CapZ, NDR, p53, HDM2, and HDM4 binding to human S100B. *Biochemistry*, 2012, 51(36):7189–7201.

[435] G. Krainer, J. Broecker, C. Vargas, J. Fanghanel, and S. Keller. Quantifying high-affinity binding of hydrophobic ligands by isothermal titration calorimetry. *Analytical Chemistry*, 2012, 84(24):10715–10722.

[436] J.P. Demers and A. Mittermaier. Binding mechanism of an SH3 domain studied by NMR and ITC. *Journal of the American Chemical Society*, 2009, 131(12):4355–4367.

[437] W.J. Wu and D.P. Raleigh. Local control of peptide conformation: Stabilization of cis proline peptide bonds by aromatic proline interactions. *Biopolymers*, 1998, 45(5):381–394.

[438] E.L. Kovrigin. NMR line shapes and multi-state binding equilibria. *Journal of Biomolecular NMR*, 2012, 53(3):257–270.

[439] E.D. Getzoff, D.E. Cabelli, C.L. Fisher, H.E. Parge, M.S. Viezzoli, L. Banci, and R.A. Hallewell. Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature*, 1992, 358(6384):347–351.

[440] S.H. Northrup and H.P. Erickson. Kinetics of protein–protein association explained by Brownian dynamics computer simulation. *Proceedings of the National Academy of Sciences of the United States of America*, 1992, 89(8):3338–3342.

[441] C. Frisch, A.R. Fersht, and G. Schreiber. Experimental assignment of the structure of the transition state for the association of barnase and barstar. *Journal of Molecular Biology*, 2001, 308(1):69–77.

[442] M. Arai, J.C. Ferreon, and P.E. Wright. Quantitative analysis of multisite protein–ligand interactions by NMR: Binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *Journal of the American Chemical Society*, 2012, 134(8):3792–3803.

[443] Y. Shen and A. Bax. Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *Journal of Biomolecular NMR*, 2007, 38(4):289–302.

[444] V. Wong and D.A. Case. Evaluating rotational diffusion from protein MD simulations. *Journal of Physical Chemistry B*, 2008, 112(19):6013–6024.

[445] J.G. de la Torre, M.L. Huertas, and B. Carrasco. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *Journal of Magnetic Resonance*, 2000, 147(1):138–146.

[446] M. Pfuhl, H.A. Chen, S.M. Kristensen, and P.C. Driscoll. NMR exchange broadening arising from specific low affinity protein self-association: Analysis of nitrogen-15 nuclear relaxation for rat CD2 domain 1. *Journal of Biomolecular NMR*, 1999, 14(4):307–320.

[447] P. Bernadó, T. Akerud, J.G. de la Torre, M. Akke, and M. Pons. Combined use of NMR relaxation measurements and hydrodynamic calculations to study protein association. Evidence for tetramers of low molecular weight protein tyrosine phosphatase in solution. *Journal of the American Chemical Society*, 2003, 125(4):916–923.

[448] Z. Liu, W.P. Zhang, Q. Xing, X.F. Ren, M.L. Liu, and C. Tang. Noncovalent dimerization of ubiquitin. *Angewandte Chemie-International Edition*, 2012, 51(2):469–472.

[449] C.D. Kroenke, J.P. Loria, L.K. Lee, M. Rance, and A.G. Palmer. Longitudinal and transverse ^{1}H–^{15}N dipolar/^{15}N chemical shift anisotropy relaxation interference: Unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *Journal of the American Chemical Society*, 1998, 120(31):7905–7915.

[450] J.B. Hall and D. Fushman. Direct measurement of the transverse and longitudinal ^{15}N chemical shift anisotropy–dipolar cross-correlation rate constants using ^{1}H-coupled HSQC spectra. *Magnetic Resonance in Chemistry*, 2003, 41(10):837–842.

[451] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R.J. Woods. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 2005, 26(16):1668–1688.

[452] A.E. Aliev and D. Courtier-Murias. Experimental verification of force fields for molecular dynamics simulations using Gly-Pro-Gly-Gly. *Journal of Physical Chemistry B*, 2010, 114(38):12358–12375.

[453] D.S. Cerutti, P.L. Freddolino, R.E. Duke, and D.A. Case. Simulations of a protein crystal with a high resolution X-ray structure: Evaluation of force fields and water models. *Journal of Physical Chemistry B*, 2010, 114(40):12811–12824.

[454] K.A. Beauchamp, Y.S. Lin, R. Das, and V.S. Pande. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *Journal of Chemical Theory and Computation*, 2012, 8(4):1409–1414.

[455] S. Piana, K. Lindorff-Larsen, and D.E. Shaw. How robust are protein folding simulations with respect to force field parameterization? *Biophysical Journal*, 2011, 100(9):L47–L49.

[456] B. Webb, A. Sali, and D. Kihara. Protein structure modeling with MODELLER. *Protein Structure Prediction, 3rd Edition*, 2014, pages 1–15.

[457] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 1983, 79(2):926–935.

[458] I. Chandrasekhar, G.M. Clore, A. Szabo, A.M. Gronenborn, and B.R. Brooks. A 500-ps molecular-dynamics simulation study of interleukin-1β in water: Correlation with nuclear magnetic resonance spectroscopy and crystallography. *Journal of Molecular Biology*, 1992, 226(1):239–250.

[459] S. Pfeiffer, D. Fushman, and D. Cowburn. Simulated and NMR-derived backbone dynamics of a protein with significant flexibility: A comparison of spectral densities for the βARK1 PH domain. *Journal of the American Chemical Society*, 2001, 123(13):3021–3036.

[460] A.J. Nederveen and A.M.J.J. Bonvin. NMR relaxation and internal dynamics of ubiquitin from a 0.2 μs MD simulation. *Journal of Chemical Theory and Computation*, 2005, 1(3):363–374.

[461] B. Han, Y.F. Liu, S.W. Ginzinger, and D.S. Wishart. SHIFTX2: Significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR*, 2011, 50(1):43–57.

[462] P.R.L. Markwick, C.F. Cervantes, B.L. Abel, E.A. Komives, M. Blackledge, and J.A. McCammon. Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. *Journal of the American Chemical Society*, 2010, 132(4):1220–1221.

[463] P. Robustelli, K.A. Stafford, and A.G. Palmer. Interpreting protein structural dynamics from NMR chemical shifts. *Journal of the American Chemical Society*, 2012, 134(14):6365–6374.

[464] J. Lehtivarjo, K. Tuppurainen, T. Hassinen, R. Laatikainen, and M. Perakyla. Combining NMR ensembles and molecular dynamics simulations provides more realistic models of protein structures in solution and leads to better chemical shift prediction. *Journal of Biomolecular NMR*, 2012, 52(3):257–267.

[465] D.W. Li and R. Bruschweiler. PPM: A side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. *Journal of Biomolecular NMR*, 2012, 54(3):257–265.

[466] S.E. Wong, K. Bernacki, and M. Jacobson. Competition between intramolecular hydrogen bonds and solvation in phosphorylated peptides: Simulations with explicit and implicit solvent. *Journal of Physical Chemistry B*, 2005, 109(11):5249–5258.

[467] B. Hess and N.F.A. van der Vegt. Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models. *Journal of Physical Chemistry B*, 2006, 110(35):17616–17626.

[468] D. Paschek, R. Day, and A.E. Garcia. Influence of water–protein hydrogen bonding on the stability of Trp-cage miniprotein. A comparison between the TIP3P and TIP4P-Ew water models. *Physical Chemistry Chemical Physics*, 2011, 13(44):19840–19847.

[469] P. Florova, P. Sklenovsky, P. Banas, and M. Otyepka. Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact. *Journal of Chemical Theory and Computation*, 2010, 6(11):3569–3579.

[470] P.S. Nerenberg and T. Head-Gordon. Optimizing protein–solvent force fields to reproduce intrinsic conformational preferences of model peptides. *Journal of Chemical Theory and Computation*, 2011, 7(4):1220–1230.

[471] H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 1987, 91(24):6269–6271.

[472] H.W. Horn, W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, G.L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *Journal of Chemical Physics*, 2004, 120(20):9665–9678.

[473] G. Lamoureux, E. Harder, I.V. Vorobyov, B. Roux, and A.D. MacKerell. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chemical Physics Letters*, 2006, 418(1-3):245–249.

[474] A.S. Woods and S. Ferre. Amazing stability of the arginine–phosphate electrostatic interaction. *Journal of Proteome Research*, 2005, 4(4):1397–1402.

[475] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G.M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J.M. Wang, and P. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 2003, 24(16):1999–2012.

[476] A.D. Mackerell, M. Feig, and C.L. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*, 2004, 25(11):1400–1415.

[477] C. Oostenbrink, A. Villa, A.E. Mark, and W.F. Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, 2004, 25(13):1656–1676.

[478] J.W. Ponder, C.J. Wu, P.Y. Ren, V.S. Pande, J.D. Chodera, M.J. Schnieders, I. Haque, D.L. Mobley, D.S. Lambrecht, R.A. DiStasio, M. Head-Gordon, G.N.I. Clark, M.E. Johnson, and T. Head-Gordon. Current status of the AMOEBA polarizable force field. *Journal of Physical Chemistry B*, 2010, 114(8):2549–2564.

[479] N. Trbovic, B. Kim, R.A. Friesner, and A.G. Palmer. Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. *Proteins: Structure, Function, and Bioinformatics*, 2008, 71(2):684–694.

[480] D. Frigyes, F. Alber, S. Pongor, and P. Carloni. Arginine–phosphate salt bridges in protein–DNA complexes: A Car-Parrinello study. *Journal of Molecular Structure (Theochem)*, 2001, 574:39–45.

[481] D. Jiao, P.A. Golubkov, T.A. Darden, and P. Ren. Calculation of protein–ligand binding free energy by using a polarizable potential. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(17):6290–6295.

[482] A.H. Ng and C.D. Snow. Polarizable protein packing. *Journal of Computational Chemistry*, 2011, 32(7):1334–1344.

[483] D. Jiao, J.J. Zhang, R.E. Duke, G.H. Li, M.J. Schnieders, and P.Y. Ren. Trypsin–ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *Journal of Computational Chemistry*, 2009, 30(11):1701–1711.

[484] T.D. Fenn, M.J. Schnieders, A.T. Brunger, and V.S. Pande. Polarizable atomic multipole X-ray refinement: Hydration geometry and application to macromolecules. *Biophysical Journal*, 2010, 98(12):2984–2992.

[485] K. Zhu, M.R. Shirts, and R.A. Friesner. Improved methods for side chain and loop predictions via the protein local optimization program: Variable dielectric model for implicitly improving the treatment of polarization effects. *Journal of Chemical Theory and Computation*, 2007, 3(6):2108–2119.

[486] I.V. Leontyev and A.A. Stuchebrukhov. Electronic continuum model for molecular dynamics simulations of biological molecules. *Journal of Chemical Theory and Computation*, 2010, 6(5):1498–1508.

[487] I. Leontyev and A. Stuchebrukhov. Accounting for electronic polarization in non-polarizable force fields. *Physical Chemistry Chemical Physics*, 2011, 13(7):2613–2626.

[488] Y. Luo and B. Roux. Simulation of osmotic pressure in concentrated aqueous salt solutions. *Journal of Physical Chemistry Letters*, 2010, 1(1):183–189.

[489] S. Berneche and B. Roux. Energetics of ion conduction through the K^+ channel. *Nature*, 2001, 414(6859):73–77.

[490] J.J. Yoo and A. Aksimentiev. Improved parametrization of Li^+, Na^+, K^+, and Mg^{2+} ions for all-atom molecular dynamics simulations of nucleic acid systems. *Journal of Physical Chemistry Letters*, 2012, 3(1):45–50.

[491] T. Simonson and P. Satpati. Simulating GTP:Mg and GDP:Mg with a simple force field: A structural and thermodynamic analysis. *Journal of Computational Chemistry*, 2013, 34(10):836–846.

[492] D.S. Cerutti, J.E. Rice, W.C. Swope, and D.A. Case. Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. *Journal of Physical Chemistry B*, 2013, 117(8):2328–2338.

[493] D.S. Cerutti, W.C. Swope, J.E. Rice, and D.A. Case. ff14ipq: A self-consistent force field for condensed-phase simulations of proteins. *Journal of Chemical Theory and Computation*, 2014, 10(10):4515–4534.

[494] R. Geney, M. Layten, R. Gomperts, V. Hornak, and C. Simmerling. Investigation of salt bridge stability in a Generalized Born solvent model. *Journal of Chemical Theory and Computation*, 2006, 2(1):115–127.

[495] A. Okur, L. Wickstrom, and C. Simmerling. Evaluation of salt bridge structure and energetics in peptides using explicit, implicit, and hybrid solvation models. *Journal of Chemical Theory and Computation*, 2008, 4(3):488–498.

[496] Y. Shang, H. Nguyen, L. Wickstrom, A. Okur, and C. Simmerling. Improving the description of salt bridge strength and geometry in a Generalized Born model. *Journal of Molecular Graphics and Modelling*, 2011, 29(5):676–684.

[497] K. Terpe. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 2003, 60(5):523–533.

[498] J.R. Huth, C.A. Bewley, B.M. Jackson, A.G. Hinnebusch, G.M. Clore, and A.M. Gronenborn. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Science*, 1997, 6(11):2359–2364.

[499] D.A. Lindhout, A. Thiessen, D. Schieve, and B.D. Sykes. High-yield expression of isotopically labeled peptides for use in NMR studies. *Protein Science*, 2003, 12(8):1786–1791.

[500] G. Wider and L. Dreier. Measuring protein concentrations by NMR spectroscopy. *Journal of the American Chemical Society*, 2006, 128(8):2571–2576.

[501] D.F. Hansen and L.E. Kay. Improved magnetization alignment schemes for spin-lock relaxation experiments. *Journal of Biomolecular NMR*, 2007, 37(4):245–255.

[502] M. Wittekind and L. Mueller. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *Journal of Magnetic Resonance*, 1993, 101(2):201–205.

[503] M. Anafi, M.K. Rosen, G.D. Gish, L.E. Kay, and T. Pawson. A potential SH3 domain-binding site in the Crk SH2 domain. *Journal of Biological Chemistry*, 1996, 271(35):21365–21374.

[504] K. Wüthrich. *NMR of Proteins and Nucleic Acids*. Wiley-Interscience, 1986.

[505] J.N. Song, K. Burrage, Z. Yuan, and T. Huber. Prediction of *cis/trans* isomerization in proteins using PSI-BLAST profiles and secondary structure information. *BMC Bioinformatics*, 2006, 7:124.

[506] Y. Xue, M.S. Pavlova, Y.E. Ryabov, B. Reif, and N.R. Skrynnikov. Methyl rotation barriers in proteins from ^{2}H relaxation data. Implications for protein structure. *Journal of the American Chemical Society*, 2007, 129(21):6827–6838.

[507] T. Bremi, R. Bruschweiler, and R.R. Ernst. A protocol for the interpretation of side-chain dynamics based on NMR relaxation: Application to phenylalanines in antamanide. *Journal of the American Chemical Society*, 1997, 119(18):4272–4284.

[508] G. Cornilescu and A. Bax. Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase. *Journal of the American Chemical Society*, 2000, 122(41):10143–10154.

[509] S. Kumar and R. Nussinov. Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophysical Journal*, 2002, 83(3):1595–1612.

[510] G.A. Jeffrey and W. Saenger. *Hydrogen Bonding in Biological Structures*. Springer-Verlalg, 1991.

[511] L. Cavallo, J. Kleinjung, and F. Fraternali. POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Research*, 2003, 31(13):3364–3366.

[512] O.W. Sørensen, G.W. Eich, M.H. Levitt, G. Bodenhausen, and R.R. Eenst. Product operator formalism for the description of NMR pulse experiments. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 1983, 16:163–192.

[513] A.D. Bain and B. Berno. Liouvillians in NMR: The direct method revisited. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2011, 59(3):223–244.

[514] H.M. McConnell. Reaction rates by nuclear magnetic resonance. *Journal of Chemical Physics*, 1958, 28(3):430–431.

[515] F.A.A. Mulder, C.A.E.M. Spronk, M. Slijper, R. Kaptein, and R. Boelens. Improved HSQC experiments for the observation of exchange broadened signals. *Journal of Biomolecular NMR*, 1996, 8(2):223–228.

APPENDICES

A Primers Design Strategies for Tandem α-spc SH3 Domains and Complete RDC Datasets Measured for All Samples

A.1 Primer Design Strategy for Constructing Tandem α-spc SH3 Domains

The original pET3d plasmid encoding the cDNA for full-length α-spc SH3 (fSH3)[225] was a gift from Prof. Reif. This plasmid was used as the starting point to construct dSH3 and another three tandem α-spc SH3 domain plasmids. The construction of dSH3 plasmid is straightforward since dSH3 can be considered as a truncated version of fSH3 (only residues 2–6 at N-terminal are deleted). The original fSH3 plasmid is used as PCR template, a pair of forward and reverse primers are designed to amplify the target cDNA region, the amplified target cDNA is finally inserted back into the original pET3d plasmid between *NcoI* and *BamHI* restriction sites.

For dSH3-*sl*-dSH3 and dSH3-*ml*-dSH3 plasmid construction, the dSH3 plasmid is used as template. Since both dSH3-*sl*-dSH3 and dSH3-*ml*-dSH3 contain several residues within the linker region (GSG and GGGGSGGGG respectively) that are not present in dSH3 plasmid, additional nucleotides corresponding to those residues have to be introduced at 5′-end of the primer. Since each tandem-SH3 protein contains two copies of identical SH3 domains, it requires two repeated cDNA regions therefore two separate PCR reactions have been carried out—one of the PCR products contains N-terminal SH3 plus half length of the linker region, whereas the other PCR product contains C-terminal SH3 and the other half linker region. These two pieces of cDNA are finally glued together and inserted back into the original pET3d plasmid between *NcoI* or *BamHI* restriction sites. Another restriction site different from *NcoI* or *BamHI* is also required to glue together these two PCR products, *BspEI* restriction site (recognition sequence TCCGGA) has been chosen in order to not introduce additional residues other than Gly or Ser, besides, *BspEI* could also work together with *NcoI* or *BamHI* for double digestion.

The construction for dSH3-*ll*-dSH3 plasmid is more challenging since its linker regions is much longer. If dSH3 plasmid is used as PCR template, too many additional

nucleotides have to be introduced at 5′-end of the primer, which may cause difficulty for primer synthesis or PCR reactions. The dSH3-*ml*-dSH3 plasmid is used as PCR template to construct dSH3-*ll*-dSH3, since it already contains nucleotides corresponding to residues GGGGSGGGG within the linker region. Two separate PCR reactions have been carried out following the strategy as mentioned earlier. The *BspEI* restriction site is still used to glue together two pieces of PCR products, however, the original *BspEI* restriction site within the linker region of dSH3-*ml*-dSH3 plasmid should be mutated, such that each PCR product only contains one single *BspEI* restriction site.

All primers that have been designed and the full amino acid sequence encoded by each plasmid are summarized as below (DNA restriction sites are marked in red, residues in disordered regions are marked in blue):

dSH3:

Forward primer:

CATGCCATGGAGCTTGTGCTAGCACTCTATG

Reverse primer:

ATCGATAAGCTTGGATCCCTAC

Full amino acid sequence:

MELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD

fSH3:

Full amino acid sequence:

MDETGKELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD

dSH3-*sl*-dSH3:

Forward primer (N domain):

CATGCCATGGAGCTTGTGCTAGCACTCTATG

Forward primer (C domain):

ACGCTCCGGAGAGCTTGTGCTAGCACTCTATG

Reverse primer (N domain):

ACGCTCCGGAACCATCTAGTTTTTTCACATAGGCAGC

Reverse primer (C domain):

ATCGATAAGCTTGGATCCCTAC

Full amino acid sequence:

MELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLDGSGELVLALYDYQE

KSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD

dSH3-*ml*-dSH3:

Forward primer (N domain):

CATGCCATGGAGCTTGTGCTAGCACTCTATG

Forward primer (C domain):

ACGCTCCGGAGGTGGTGGTGAGCTTGTGCTAGCACTCTATG

Reverse primer (N domain):

ACGCTCCGGAACCACCACCACCATCTAGTTTTTTCACATAGGCAGC

Reverse primer (C domain):

ATCGATAAGCTTGGATCCCTAC

Full amino acid sequence:

MELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLDGGGGSGGGGELVLA

LYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD

dSH3-*ll*-dSH3:

Forward primer (N domain):

CATGCCATGGAGCTTGTGCTAGCACTCTATG

Forward primer (C domain):

ACGCTCCGGAGGCGGAGGCTCAGGTGGTGGTGGTTCTGGAGGTGGTGGTGAGCTTG

Reverse primer (N domain):

ACGCTCCGGACCCGCCTCCGCCTGAACCACCACCACCGGAACCACCACCACCATCTAG

Reverse primer (C domain):

ATCGATAAGCTTGGATCCCTAC

Full amino acid sequence:

MELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLDGGGGSGGGGSGGGG

SGGGGSGGGGSGGGGELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD

A.2 Complete RDC Datasets Measured in 5% PEG/hexanol ($r = 0.85$)

Experimentally measured RDCs for various constructs of α-spc SH3 domain at pH 3.5, 20 mM sodium citrate, 5% PEG/hexanol ($r = 0.85$) condition are shown in Table A.1. The high salt sample contains additionally 100 mM NaCl. For α-spc SH3 tandems, each value represents the average between the N- and C-terminal sites. Abbreviations are: (w) weak peak, (o) overlapped peak, (p) proline residue, (s) misshaped peak.

Table A.1.: Experimentally measured RDCs in 5 % PEG/hexanol ($r = 0.85$)

Residue	dSH3	dSH3	fSH3	fSH3 (high salt)	dSH3-*sl*-dSH3	dSH3-*sl*-dSH3	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3	dSH3-*ll*-dSH3
9	w	−32.4	−9.4	−20.0	12.5	12.0	2.8	−0.2	−0.5
10	w	w	−16.0	w	15.4	14.6	7.8	−0.5	−1.3
11	4.6	5.1	2.5	5.5	15.8	12.5	7.0	8.5	9.7
12	0.5	2.5	0.2	w	w	w	w	−2.8	−4.1
13	w	s	−2.3	0.0	9.9	9.3	6.7	5.7	6.5
14	−11.9	−18.0	−0.5	5.2	15.0	14.8	9.5	7.7	8.9
15	w	w	11.7	17.8	15.1	14.9	9.8	11.9	13.9
16	14.6	11.8	10.4	16.8	22.9	22.1	10.8	13.3	15.4
17	w	34.2	12.5	21.9	1.1	1.3	1.5	8.3	8.9
18	w	w	−25.0	w	4.3	4.0	−1.1	−10.0	−11.8
19	w	w	19.7	32.4	−23.0	−22.4	−9.9	−4.0	−5.1
20	p	p	p	p	p	p	p	p	p
21	32.4	w	15.3	24.8	4.5	4.8	4.9	7.3	8.9
22	−4.6	−2.0	−10.7	−20.4	−13.0	−13.2	−6.8	−11.6	−13.0

Table A.1.: continued

Residue	dSH3	dSH3	fSH3	fSH3 (high salt)	dSH3-*sl*-dSH3	dSH3-*sl*-dSH3	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3	dSH3-*ll*-dSH3
23	w	w	−7.2	−7.0	16.3	16.1	6.3	3.0	2.6
24	−6.6	−16.2	1.9	2.8	19.3	18.6	10.1	9.8	11.1
25	25.1	33.4	14.1	w	14.1	13.7	8.4	11.1	12.7
26	25.1	24.3	8.8	w	13.3	12.4	9.1	5.8	6.5
27	5.7	6.7	−5.6	−10.3	4.0	3.1	1.9	−2.1	−3.4
28	w	w	−14.7	w	2.3	3.4	−0.7	−3.0	−3.0
29	36.7	o	17.0	25.9	−15.9	−15.4	o	1.3	2.0
30	−19.9	−22.8	−9.2	w	o	o	o	3.0	3.2
31	−16.3	−23.5	−5.6	−11.9	16.7	16.1	7.4	7.1	4.8
32	−26.3	−22.2	−9.4	−21.2	−2.7	−3.5	−3.9	−5.4	−6.2
33	w	w	13.5	w	−26.1	−26.5	w	−2.0	−2.0
34	19.7	17.9	−6.1	w	4.3	7.0	3.0	−0.6	−1.5
35	w	w	−21.8	w	10.1	9.8	0.6	−2.3	−2.9
36	−15.0	−20.0	−14.1	−23.4	10.1	9.6	2.7	−2.1	−3.6
37	−11.7	−20.0	0.2	−2.6	18.4	17.5	9.3	6.9	8.4
38	−14.2	−14.5	−7.9	−13.7	9.5	9.2	4.0	2.7	2.9
39	8.1	o	3.2	o	−17.2	−15.6	o	−2.8	−2.7
40	5.4	4.8	o	o	−12.4	−12.1	−5.3	−0.4	−0.1
41	−35.2	w	−20.2	w	−0.2	0.1	−2.8	−6.7	−7.7
42	w	w	−23.7	w	7.4	7.5	3.0	−4.1	−4.8
43	w	w	o	w	5.2	5.4	−1.3	−10.3	−13.1
44	8.0	w	−1.2	−4.5	−13.6	−13.0	−9.8	−9.9	−11.6
45	o	o	6.6	o	o	o	o	o	o
46	s	s	o	w	−26.9	−26.1	−13.3	−8.5	−9.6
47	28.7	31.8	12.2	21.2	−21.6	−21.5	−11.3	−0.7	−1.9
48	−12.1	−10.8	−9.6	−17.5	−5.7	−5.8	−5.9	−7.7	−8.7
49	9.2	8.1	12.9	24.9	o	o	o	−0.6	−2.1
50	8.4	8.3	10.0	16.6	o	o	o	o	o

Table A.1.: continued

Residue	dSH3	dSH3	fSH3	fSH3 (high salt)	dSH3-*sl*-dSH3	dSH3-*sl*-dSH3	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3	dSH3-*ll*-dSH3
51	17.8	26.4	9.0	13.2	−23.9	−23.1	−11.9	−6.3	−6.5
52	26.4	30.0	10.2	17.6	−22.7	−22.6	−9.7	−6.8	−8.7
53	−34.5	−36.4	−27.4	w	4.8	4.7	−0.4	−9.4	−11.9
54	p	p	p	p	p	p	p	p	p
55	w	w	15.5	w	−27.4	−25.7	−11.1	−4.9	−6.3
56	40.9	w	23.4	41.5	−29.7	−29.7	−11.6	−4.7	−6.4
57	w	w	12.7	w	−8.0	−8.5	−3.1	2.6	3.3
58	s	s	−8.8	−13.9	−2.9	−2.1	−3.7	−2.1	−2.4
59	o	o	−11.8	o	o	o	o	o	o
60	0.1	−0.2	−0.3	−1.3	12.1	11.9	5.3	6.3	6.8

A.3 Complete RDC Datasets Measured in 5% PEG/hexanol ($r = 0.96$)

Experimentally measured RDCs for various constructs of α-spc SH3 domain at pH 3.5, 20 mM sodium citrate, 5% PEG/hexanol ($r = 0.96$) condition are shown in Table A.2. For α-spc SH3 tandems, each value represents the average between the N- and C-terminal sites. Abbreviations are: (w) weak peak, (o) overlapped peak, (p) proline residue.

Table A.2.: Experimentally measured RDCs in 5 % PEG/hexanol ($r = 0.96$)

Residue	dSH3	fSH3	dSH3-*sl*-dSH3	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3
9	−22.3	−5.2	12.6	4.2	−0.6
10	−27.9	−8.4	14.4	11.1	−0.8
11	−0.5	1.4	12.5	7.7	5.5
12	2.2	0.1	w	w	−2.1
13	−16.4	−0.8	9.4	6.0	3.7

Table A.2.: continued

Residue	dSH3	fSH3	dSH3-*sl*-dSH3	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3
14	−12.1	0.3	14.4	9.7	5.0
15	15.9	7.0	14.3	10.7	7.5
16	8.9	6.2	22.1	12.9	8.4
17	24.5	7.0	1.4	2.3	5.8
18	w	−14.7	4.8	−1.0	−7.4
19	43.4	10.1	−22.7	−12.0	−1.7
20	p	p	p	p	p
21	30.0	8.4	4.1	5.7	5.1
22	−1.7	−6.3	−12.7	−6.6	−7.7
23	−19.0	−4.5	15.8	7.2	1.4
24	−8.5	0.4	17.9	10.5	5.7
25	25.9	7.7	13.2	10.4	6.9
26	17.9	5.4	11.0	8.5	4.0
27	4.9	−2.9	3.3	1.5	−2.1
28	−19.5	−8.5	3.9	−0.1	−1.9
29	39.3	9.0	−15.1	o	0.7
30	−17.7	−5.1	o	o	1.6
31	−19.9	−3.2	14.4	8.3	4.5
32	−16.0	−5.5	−3.3	−4.2	−4.8
33	35.7	6.5	−26.5	−10.2	−1.1
34	2.9	−3.9	w	2.6	−0.2
35	−30.2	−11.8	9.3	3.4	−1.5
36	−11.5	−7.8	8.6	2.3	−2.3
37	−21.4	0.5	17.8	8.8	4.6
38	−11.4	−3.9	9.3	3.5	1.5
39	9.0	1.5	−16.6	o	−0.7
40	4.1	o	−11.0	−4.0	0.5
41	−28.0	−11.0	−0.4	−3.1	−4.5
42	w	−13.0	7.5	1.7	−3.1
43	w	−16.6	5.2	−1.6	−7.5
44	16.9	−1.1	−13.4	−7.9	−6.9

Table A.2.: continued

Residue	dSH3	fSH3	dSH3-*sl*-dSH3	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3
45	o	2.9	o	o	o
46	20.1	8.4	−27.3	−15.8	−5.1
47	22.9	6.1	−19.7	−11.1	−2.7
48	−8.3	−6.2	−6.4	−6.1	−4.4
49	6.5	6.9	o	o	−0.1
50	6.0	5.2	o	o	o
51	20.6	4.2	−23.1	−12.4	−3.7
52	21.5	5.2	−20.1	−9.3	−4.3
53	−26.6	−14.8	4.3	−1.3	−6.7
54	p	p	p	p	p
55	30.1	7.8	−26.9	−19.1	−1.4
56	46.4	12.6	−29.3	−15.2	−2.0
57	7.6	6.8	−8.3	−2.5	2.4
58	−8.8	−5.0	−2.3	−3.6	−1.6
59	o	−6.3	o	o	o
60	−0.2	0.2	11.8	6.5	3.6

B Numerical Simulation for PD-CPMG Pulse Sequence

B.1 Basic Principles of Numerical Simulation for Spin Dynamics

Numerical simulation for spin dynamics is one of the most important tools for NMR pulse sequence development[134]. Since NMR experiments are usually carried out on samples containing large number of molecules ($\sim 10^{23}$), it is necessary to describe the property of the whole system based on statistical approaches rather than focus on each single molecule. The information about quantum systems containing large numbers of particles is encoded in the density operator $\hat{\rho}$ (or density matrix ρ since it is usually expressed in matrix form), which is similar to the fact that all information about single particle quantum system is contained in the wavefunction $|\Psi\rangle$. The relationship between density operator $\hat{\rho}$ and wavefunction $|\Psi\rangle$ is:

$$\hat{\rho} = \overline{|\Psi\rangle\langle\Psi|} \tag{B.1}$$

the overbar indicates ensemble averaging. For NMR related studies, only those nuclear spin degrees of freedom need to be taken into account, therefore $\hat{\rho}$ is usually called spin density operator.

In order to learn about the evolution of the system during each NMR pulse sequence, it is necessary to evaluate the time evolution of spin density operator $\hat{\rho}$, which can be described by Liouville–von Neumann Equation (B.2):

$$\frac{d}{dt}\hat{\rho}(t) = -i[\hat{H}(t), \hat{\rho}(t)] \tag{B.2}$$

$\hat{H}$ is Hamiltonian of the whole system. This equation is simply based on the Schrödinger equation which describes the time evolution of single particle wavefunction $|\Psi\rangle$.

The solution to Equation (B.2) can be formally written as:

$$\hat{\rho}(t) = \exp(-i\hat{H}t)\hat{\rho}(0)\exp(i\hat{H}t) \tag{B.3}$$

It should be noted that the Equation (B.2) is defined in Hilbert space, since the basis set is usually chosen as the set of eigenvectors for single particle wavefunctions $|\Psi\rangle$. The Liouville–von Neumann equation can also be expressed in Liouville space (i.e., the superspace of Hilbert space):

$$\frac{d}{dt}\hat{\rho} = -i\hat{\hat{L}}\hat{\rho} \tag{B.4}$$

The formal solution to Equation (B.4) can be expressed as:

$$\hat{\rho}(t) = \exp(-i\hat{\hat{L}}t)\hat{\rho}(0) \tag{B.5}$$

$\hat{\hat{L}}$ is an operator in Liouville space called Liouvillian, which can also be regarded as a superoperator in Hilbert space according to its definition:

$$\hat{\hat{L}}(t) = [\hat{H}(t),\ ] \tag{B.6}$$

According to Equation (B.2) or (B.4), it is necessary to learn about Hamiltonian $\hat{H}$ (or Liouvillian $\hat{\hat{L}}$) of the system in order to evaluate the time evolution of spin density operator $\hat{\rho}$. The calculation based on Liouville–von Neumann Equation usually requires explicit matrix manipulations, therefore it becomes tedious when the system contains large numbers of different spin species. Some alternative strategies have been developed to simplify such calculations, for example, product operator formalism is another commonly used approach[512]. This method allows for straightforward evaluation of the time evolution of spin density operator $\hat{\rho}$ without explicit matrix treatment, however, it can only be utilized when Hamiltonian $\hat{H}$ of the system is relatively simple (such as only containing J-couplings and chemical shifts). As Hamiltonian $\hat{H}$ of the system becomes more complicated, it is still necessary to resort to explicit matrix calculation to evaluate the time evolution of spin density operator $\hat{\rho}$.

The two different forms of Liouville–von Neumann Equations (Equations (B.2) and (B.4)) are equivalent except that they are expressed in different spaces[513], therefore the results as obtained from these two forms of equations should be identical. However, there are more advantages for performing numerical calculation in Liouville space—the eigenstates in Liouville space usually correspond to different terms of coherences in spin density operator $\hat{\rho}$, which are directly related to different types of observables, besides, the time evolution of spin density operator $\hat{\rho}$ can be described in Liouville space more conveniently as relaxation effect is also taken into account. The Liouville–von Neumann Equation in the presence of relaxation effects is:

$$\frac{d}{dt}\hat{\rho} = -i\hat{\hat{L}}\hat{\rho} - \hat{\hat{R}}(\hat{\rho} - \hat{\rho}^{eq}) \tag{B.7}$$

$\hat{\hat{R}}$ is the relaxation superoperator that includes all the relaxation effects. $\hat{\rho}^{eq}$ is the equilibrium spin density operator (k_B is Boltzmann constant):

$$\hat{\rho}^{eq} = \frac{e^{-\hat{H}/k_BT}}{Tr\{e^{-\hat{H}/k_BT}\}} \tag{B.8}$$

Since most NMR experiments for biomolecules are carried out at near room temperature ($\sim$ 300 K), the high-temperature approximation is always valid such that $\hat{\rho}^{eq}$ can be simplified as:

$$\begin{aligned}
\hat{\rho}^{eq} &= \frac{e^{-\hat{H}/k_BT}}{Tr\{e^{-\hat{H}/k_BT}\}} \\
&\approx \frac{\hat{E} - \hat{H}/k_BT}{Tr\{\hat{E} - \hat{H}/k_BT\}} \\
&\approx \frac{\hat{E} - \hat{H}/k_BT}{Tr\{\hat{E}\}} \\
&\approx \hat{E}/N - \hat{H}/(Nk_BT)
\end{aligned} \tag{B.9}$$

N is the dimensionality of Hilbert space and is equal to 2^M (M is the number of spin-1/2 nuclei). $\hat{E}$ is the identity operator in Hilbert space.

B.2 Numerical Simulation Setup for PD-CPMG

B.2.1 Two-spin System

The setup for two-spin system numerical simulation is mainly based on the properties of ($^1H^N$,^{15}N) spin system from protein backbone amide group. Since ^1H and ^{15}N are both spin-1/2 nuclei, the number of bases in Hilbert space is $2^2 = 4$, such that the spin density operator $\hat{\rho}$ and Hamiltonian $\hat{H}$ are both expressed as 4×4 matrices in Hilbert space. In Liouville space, the density operator $\hat{\rho}$ is rewritten as 16×1 vector, the Liouvillian $\hat{\hat{L}}$ is 16×16 matrix (corresponding to the size of superoperators in Hilbert space), the relaxation superoperator $\hat{\hat{R}}$ is also expressed as 16×16 matrix.

In this work, all numerical simulations have been carried out in Liouville space in order to take relaxation effects into account more conveniently (Equation (B.7)). Since the identity matrix E does not evolve with time, it can be excluded from the basis set such that $\hat{\rho}$ is expressed as 15×1 vector, $\hat{\hat{L}}$ and $\hat{\hat{R}}$ are both expressed as 15×15 matrices. However, it should be noted that in the case of $\hat{\rho}^{eq} \neq 0$, the identity matrix E should still be included in order to express the time evolution differential equation for $\hat{\rho}$ as homogeneous form[139]. In most real pulse sequences (such as PD-CPMG), the phase cycle effectively cancels out all equilibrium magnetizations such that $\hat{\rho}^{eq} = 0$, therefore E can be excluded. The basis set used in the numerical simulation are constructed based on shift bases (such as H_+ and H_-) instead of Cartesian bases (such as H_x and H_y). Each shift basis has characteristic Larmor precession frequency, therefore relaxation effects between different bases can be evaluated conveniently. The total 15 bases that have been included for simulating two-spin system are:

$$\begin{aligned}\{&N_+, N_-, N_z, N_+H_+, N_-H_+, N_zH_+, H_+,\\ &N_+H_-, N_-H_-, N_zH_-, H_-, N_+H_z, N_-H_z, N_zH_z, H_z\}\end{aligned} \tag{B.10}$$

The numerical simulation focuses only on the relaxation delay τ_{rel} period, since ^{15}N magnetization relaxation is measured by setting different values of τ_{rel} in PD-

CPMG. The density operator at the beginning is $\hat{\rho}(0) = \mathrm{N_x}$ $(= (\mathrm{N_+} + \mathrm{N_-})/2)$. The Liouvillian $\hat{\hat{L}}$ includes all coherent evolution effects such as J-couplings, *rf* pulses and chemical shifts. The relaxation superoperator $\hat{\hat{R}}$ includes relaxation effects caused by DD and CSA interactions, as well as chemical exchange effects. Both $\hat{\hat{L}}$ and $\hat{\hat{R}}$ are calculated as the symbolic forms in Maple first, then numerical values are inserted to generate $\hat{\hat{L}}$ and $\hat{\hat{R}}$ as numerical matrices for explicit matrix calculation in Python.

In PD-CPMG pulse sequence, ^{1}H channel is continuously applied with *rf* pulses for ^{1}H decoupling (DIPSI-2 or other decoupling sequences) during τ_{rel}, and at any time point the decoupling *rf* pulse has phase along x or $-$x. For ^{15}N channel, there are four different states: (i) no *rf* pulses applied; (ii) *rf* pulse along x; (iii) *rf* pulse along y; (iv) *rf* pulse along $-$y. Therefore there are totally $2 \times 4 = 8$ forms of Liouvillian $\hat{\hat{L}}$ when considering both ^{1}H and ^{15}N channels. Since $\hat{\hat{R}}$ always remains the same, the whole time period of τ_{rel} can be divided into many time intervals, such that during each interval $\hat{\hat{L}}$ remains the same. The time evolution of spin density operator $\hat{\rho}$ is calculated within each individual time interval τ: since both $\hat{\hat{L}}$ and $\hat{\hat{R}}$ remain constant during τ, the evolution of spin density operator $\hat{\rho}$ can be expressed as (assume that $\hat{\rho}^{eq} = 0$):

$$\hat{\rho}(t+\tau) = \exp(-i\hat{\hat{L}}\tau - \hat{\hat{R}}\tau)\hat{\rho}(t) = \exp(\hat{\hat{K}}\tau)\hat{\rho}(t) \tag{B.11}$$

therefore it is only necessary to evaluate $\exp(\hat{\hat{K}}\tau)$ $(\hat{\hat{K}} = -i\hat{\hat{L}} - \hat{\hat{R}})$. However, in general $[\hat{\hat{L}}, \hat{\hat{R}}] \neq 0$ such that $\exp(\hat{\hat{K}}\tau) \neq \exp(-i\hat{\hat{L}}\tau)\exp(-\hat{\hat{R}}\tau)$, besides, $\hat{\hat{K}}$ is usually not a unitary matrix therefore it cannot be diagonalized. In practice, the numerical calculation for $\exp(\hat{\hat{K}}\tau)$ is based on the following relationship ($\hat{\hat{E}}$ is the identity matrix in Liouville space):

$$\exp(\hat{\hat{K}}\tau) = \lim_{\varepsilon \to 0} (\hat{\hat{E}} + \hat{\hat{K}}\varepsilon)^{\frac{\tau}{\varepsilon}} \tag{B.12}$$

the results of $\exp(\hat{\hat{K}}\tau)$ is independent on the choice of ε as long as $\varepsilon \ll \tau$.

The numerical simulation has been carried out for an array of τ_{rel} corresponding to experimental values, N_x is evaluated at the end of each τ_{rel}, then R_2^{app} is extracted from exponential curve fitting.

In order to take solvent exchange effect into account, it is assumed that all multiple quantum coherences involving labile ^1H (such as N_+H_+ and N_+H_z) have additional relaxation rates as k_{ex}[327]. The solvent exchange effects for H_+, H_- and H_z terms on labile ^1H can only be correctly modeled as water magnetization is also included into the simulation[327]. However, since the initial spin density operator is N_x and very small amount of H_+, H_- or H_z have been generated during τ_{rel}, it causes little errors by neglecting water magnetizations, therefore for simplicity it is just assumed that those single-quantum coherences on labile ^1H (such as H_+, H_- and H_z) also have additional relaxation rates as k_{ex}.

In summary, for two-spin system the Liouvillian $\hat{\hat{L}}$ includes all contributions to the coherent evolution:

(i) J-couplings: $^1J_{NH}$ ($= -93$ Hz)

(ii) *rf* pulses: $\omega_1(H^N)$, $\omega_1(N)$

(iii) chemical shifts: $\omega(H^N)$, $\omega(N)$

The relaxation superoperator $\hat{\hat{R}}$ includes all the relaxation effects:

(i) DD relaxation on H^N, N (due to DD interaction between H^N and N)

(ii) CSA relaxation on H^N, N

(iii) DD/CSA cross-correlation on H^N, N

(iv) Solvent exchange effect (k_{ex}) on H^N

B.2.2 Three-spin System

The setup for three-spin system numerical simulation is mainly based on the properties of ($^1H^N$,^{15}N) spin system from protein backbone amide group, plus an

additional $^1H^\alpha$ spin within the same amino acid residue. The number of bases in Liouville space is $4^3 - 1 = 63$ (still assume $\hat{\rho}^{eq} = 0$ such that E can be excluded from the bases), therefore $\hat{\rho}$ is represented by 63×1 vector, $\hat{\hat{L}}$ and $\hat{\hat{R}}$ are both represented by 63×63 matrices. The basic strategy of numerical calculation is the same as used for two-spin system.

In summary, for three-spin system the Liouvillian $\hat{\hat{L}}$ includes all contributions to the coherent evolution:

(i) J-couplings: $^1J_{NH}$ ($= -93$ Hz), $^2J_{NH\alpha}$ ($= 5$ Hz), $^3J_{HH\alpha}$ ($= 10$ Hz)

(ii) *rf* pulses: $\omega_1(H^N)$, $\omega_1(N)$, $\omega_1(H^\alpha)$

(iii) chemical shifts: $\omega(H^N)$, $\omega(N)$, $\omega(H^\alpha)$,

The relaxation superoperator $\hat{\hat{R}}$ includes all the relaxation effects:

(i) DD relaxation on H^N, N, H^α

(ii) CSA relaxation on H^N, N

(iii) DD/CSA cross-correlation on H^N, N

(iv) DD/DD cross-correlation on H^N, N, H^α

(v) Solvent exchange effect (k_{ex}) on H^N

C Numerical Simulation for CP-HISQC Pulse Sequence

C.1 Spin Dynamics Simulation in the Presence of Chemical Exchange Process

The spin dynamics simulation for chemical exchange process can be carried out in similar way as for other types of relaxation mechanisms (such as DD and CSA), since all relaxation processes have the common feature that they conduct incoherent magnetization transfers through stochastic processes. The Liouville equation that takes relaxation effects into account (Equation (B.7)) can still be used to calculate spin density operator $\hat{\rho}$ evolution in the presence of chemical exchange effects.

Chemical exchange process is different from many other types of relaxation mechanisms (such as DD and CSA) in that it usually happens on much slower timescale. The time modulations of DD and CSA interactions are determined by molecular motions on ps~ns timescale, which is much faster than the timescale of nuclear spin relaxation (i.e., T_1 or T_2), therefore relaxation effects caused by DD or CSA interactions should be evaluated according to the Redfield theory[34]. Chemical exchange process is characterized by the time constant τ_{ex} ($= 1/k_{ex}$) which is usually on ~ms timescale, therefore it contributes to relaxation mainly through the spectral density component $J(0)$, since the high-frequency spectral density component $J(\omega)$ is usually negligible. The effects of chemical exchange process usually can be incorporated into the relaxation superoperator $\hat{\hat{R}}$ based on k_{ex}, for example, the most simple chemical exchange model between two different magnetization species (A and B) can be described by Bloch–McConnell equation[514] (k_1, k_2 are exchange rate constants of the two magnetizations):

$$\begin{aligned} \frac{dM_A}{dt} &= -k_1 M_A + k_2 M_B \\ \frac{dM_B}{dt} &= +k_1 M_A - k_2 M_B \end{aligned} \tag{C.1}$$

The chemical exchange process can be further divided into several groups depending on its detailed mechanism, such as conformational exchange or solvent exchange

processes. The conformational exchange process is usually caused by the change of molecular conformation, and it also includes some other effects such as proline *cis–trans* isomerization, the common feature is that such process always happens within the same molecule. The solvent exchange effects arise from the exchange of nuclear spin magnetizations between labile protons (such as amide group $^1H^N$) and the solvent (such as water $^1H^{water}$), therefore such process always happens between different molecules. One of the most direct consequences of chemical exchange process is the additional relaxation rate R_{ex} caused by that, which may induce additional signal loss during each coherence transfer step in NMR pulse sequence. It has been shown that R_{ex} caused by conformational exchange can be largely suppressed by performing frequent pulsing, and this idea forms the basis of certain experiments such as CPMG-HSQC[515]. However, R_{ex} caused by solvent exchange process cannot be suppressed by such strategy, therefore it poses a major challenge for studying labile ^1H on disordered proteins under physiological conditions, which could significantly undermine signal-to-noise of NMR spectra.

The spin dynamics simulation in the presence of solvent exchange process can be carried out by including solvent exchange effects into the relaxation superoperator $\hat{\hat{R}}$. For multi-quantum coherences involved with labile ^1H (such as N_+H_+ and N_+H_z), the effect is simply to increase their relaxation rates by the amount of k_{ex}, which can also be regarded as another type of second kind scalar relaxation[140,349]. For single quantum coherences on labile ^1H (i.e., H_+, H_- and H_z), since solvent exchange process conducts magnetization exchange with water, it is necessary to also include water magnetization modes (i.e., H_+^{water}, H_-^{water} and H_z^{water}) into spin dynamics simulation[327]. The modified form of Bloch–McConnell equation can be described as follows:

$$\begin{aligned} \frac{dM_A}{dt} &= -k_{ex}(1-f)M_A + k_{ex}fM_B \\ \frac{dM_B}{dt} &= +k_{ex}(1-f)M_A - k_{ex}fM_B \end{aligned} \tag{C.2}$$

M_A and M_B represents ^{1}H magnetizations on amide group and water respectively (such as H_z^N and H_z^{water}), $f = M_A/(M_A + M_B)$, k_{ex} is the solvent exchange rate constant. It can be verified that the equilibrium amide group and water ^{1}H magnetizations have been correctly preserved in Equation (C.2).

C.2 Numerical Simulation Setup for CP-HISQC

The setup of numerical simulation for CP-HISQC is mainly based on the properties of ($^1H^N$,^{15}N) spin system from protein backbone amide group, besides, those modes from water magnetization $^1H^{water}$ have also been included to correctly model solvent exchange effects. The number of bases in Liouville space is $4^2 + 3 - 1 = 18$ (assume $\hat{\rho}^{eq} = 0$ such that E can be excluded from the bases). $\hat{\rho}$ is represented by 18×1 vector, $\hat{\hat{L}}$ and $\hat{\hat{R}}$ are both represented by 18×18 matrices. The total 18 bases that have been included for numerical simulations are:

$$\begin{aligned}\{&N_+, N_-, N_z, N_+H_+, N_-H_+, N_zH_+, H_+,\\ &N_+H_-, N_-H_-, N_zH_-, H_-, N_+H_z, N_-H_z, N_zH_z, H_z,\\ &H_+^{water}, H_-^{water}, H_z^{water}\}\end{aligned} \tag{C.3}$$

The numerical simulation for CP-HISQC mainly focuses on the cross-polarization period τ_{mix} at the beginning part, since it is the key element for CP-HISQC and provides much higher coherence transfer efficiency that many other schemes (such as INEPT). During τ_{mix} time period, continuous *rf* pulses are applied on both ^{1}H and ^{15}N channels, besides, *rf* pulses have alternating phases along x or $-$x, and they are synchronous on ^{1}H and ^{15}N channels as required by DIPSI-2. Therefore τ_{mix} can be divided into a series of time intervals, such that *rf* pulses have fixed phase during each time interval τ. The spin density operator at the beginning is $\hat{\rho}(0) = H_x(= (H_+ + H_-)/2)$, the magnitude of N_x is evaluated at the end of τ_{mix}. Within each time interval τ, since both $\hat{\hat{L}}$ and $\hat{\hat{R}}$ remain constant, the evolution of spin density operator $\hat{\rho}$ can still be described by Equation (B.11), and in practice is calculated with Equation (B.12).

In order to focus on the effects of solvent exchange on spin density operator $\hat{\rho}$ evolution, the relaxation superoperator $\hat{\hat{R}}$ has ignored DD or CSA relaxation mechanisms (also because τ_{mix} is usually very short, $\sim$ 10 ms, therefore DD or CSA relaxation causes very little signal loss during τ_{mix}).

In summary, the Liouvillian $\hat{\hat{L}}$ includes all contributions to the coherent evolution:

(i) J-couplings: $^{1}\mathrm{J}_{\mathrm{NH}}$ ($= -93$ Hz)

(ii) *rf* pulses: $\omega_1(\mathrm{H}^{\mathrm{N}})$, $\omega_1(\mathrm{N})$, $\omega_1(\mathrm{H}^{\mathrm{water}})$

(iii) chemical shifts: $\omega(\mathrm{H}^{\mathrm{N}})$, $\omega(\mathrm{N})$, $\omega(\mathrm{H}^{\mathrm{water}})$

The relaxation superoperator $\hat{\hat{R}}$ only includes solvent exchange effects characterized by k_{ex}, all other types of relaxation mechanisms such as DD, CSA and potential radiation damping on water have been neglected.

D Details about ^{15}N Relaxation Parameters Calculation from MD Trajectories

D.1 Backgrounds about ^{15}N Relaxation Parameters and the Relationship with Protein Dynamics on ps~ns Timescale

Nuclear spin relaxation is one of the most important tools for characterizing protein dynamics[50]. Detailed information about protein ps~ns timescale dynamics can be extracted from the analysis of backbone ^{15}N relaxation parameters such as R_1, R_2 and ^{15}N–{^{1}H} NOE. The major advantage of analyzing ^{15}N relaxation parameters is due to the relatively simple local chemical environment of protein backbone amide group, besides, ^{15}N-labeling is one of the most economical isotope labeling strategies for protein expression.

The ^{15}N relaxation on protein backbone amide group is usually dominated by DD and CSA interactions ($\hat{H}_{DD}$ and $\hat{H}_{CSA}$), both types of interactions are anisotropic in nature and can be expressed as the general form of interaction Hamiltonian $\hat{H}_{int}$:

$$\hat{H}_{int} = \left(\hat{I}_x, \hat{I}_y, \hat{I}_z\right) \begin{pmatrix} A_{xx} & A_{xy} & A_{xz} \\ A_{yx} & A_{yy} & A_{yz} \\ A_{zx} & A_{zy} & A_{zz} \end{pmatrix} \begin{pmatrix} \hat{X}_x \\ \hat{X}_y \\ \hat{X}_z \end{pmatrix} = \hat{\mathbf{I}} \cdot \mathbf{A} \cdot \hat{\mathbf{X}} \tag{D.1}$$

$\hat{\mathbf{I}}$ is nuclear spin operator, $\hat{\mathbf{X}}$ is magnetic field vector ($\mathbf{B}_0$) or another nuclear spin operator. $\mathbf{A}$ is the interaction tensor that reflects the anisotropy of local interactions, besides, $\mathbf{A}$ is usually second-rank tensor for most types of interactions related to NMR studies (such as DD and CSA interactions).

Under solution state conditions, molecules undergo both overall tumbling motion (~ns timescale) and fast internal motion (~ps timescale), which cause fast modulation of $\hat{H}_{int}$, therefore it creates random fluctuating local magnetic field that contributes to nuclear spin relaxation. Since the timescale of those stochastic motions (ps~ns) is usually much shorter than the timescale of ^{15}N magnetization relaxation (i.e., T_1 or T_2), the Redfield theory[34] should be utilized to evaluate the effects of random fluctuating field on nuclear spin relaxation.

Such random fluctuating field is stochastic in nature, one of the most commonly used tools for characterizing such effects is the correlation function $G(\tau)$, which describes the correlation of stochastic effects (such as $\hat{H}_{int}$) at two different time points with interval τ:

$$G(\tau) = \langle H_{int}(t) H_{int}(t+\tau) \rangle \tag{D.2}$$

besides, it is usually assumed that correlation function has exponential decay form:

$$G(\tau) = G(0) \exp(-|\tau|/\tau_c) = G(0) g(\tau) \tag{D.3}$$

τ_c is the correlation time, $G(0) = \langle H_{int}^2(0) \rangle$ reflects the absolute magnitude of stochastic interactions.

The spectral density function $J(\omega)$ reflects the power distribution at different frequencies of the random field created by stochastic process, and it can be obtained by performing Fourier transform to the correlation function $g(\tau)$:

$$J(\omega) = \int_{-\infty}^{\infty} g(\tau) \exp(-i\omega\tau) d\tau \tag{D.4}$$

In the case of exponential decay form $g(\tau) = \exp(-\tau/\tau_c)$, $J(\omega)$ can be expressed as:

$$J(\omega) = \frac{2\tau_c}{5(1+\omega^2\tau_c^2)} \tag{D.5}$$

which has Lorentzian functional form.

According to calculation results based on Redfield theory, the relationship between ^{15}N relaxation parameters and spectral density function $J(\omega)$ can be expressed as:

$$\begin{aligned}
\mathrm{R_1} &= \frac{d^2}{4}[J(\omega_\mathrm{H}-\omega_\mathrm{N})+3J(\omega_\mathrm{N})+6J(\omega_\mathrm{H}+\omega_\mathrm{N})]+3c^2J(\omega_\mathrm{N}) && \text{(D.6)}\\
\mathrm{R_2} &= \frac{d^2}{8}[4J(0)+J(\omega_\mathrm{H}-\omega_\mathrm{N})+3J(\omega_\mathrm{N})+6J(\omega_\mathrm{H})+6J(\omega_\mathrm{H}+\omega_\mathrm{N})] \\
&+ \frac{c^2}{2}[4J(0)+3J(\omega_\mathrm{N})]+\mathrm{R_{ex}} && \text{(D.7)}\\
\mathrm{NOE} &= 1+\frac{\gamma_\mathrm{N}}{\mathrm{R_1}}\frac{d^2}{4}[6J(\omega_\mathrm{H}+\omega_\mathrm{N})-J(\omega_\mathrm{H}-\omega_\mathrm{N})] && \text{(D.8)}\\
\eta_\mathrm{xy} &= \frac{dc}{2}[4J(0)+3J(\omega_\mathrm{N})]\left(\frac{3\cos^2\beta-1}{2}\right) && \text{(D.9)}
\end{aligned}$$

the dipolar and CSA constants are $d=\gamma_\mathrm{N}\gamma_\mathrm{H}\mu_0\hbar/(4\pi r_\mathrm{HN}^3)$ and $c=(\omega_\mathrm{N}\Delta\sigma)/3$, respectively. $\Delta\sigma$ is the magnitude of CSA on ^{15}N, β is the polar angle of N–H bond vector in the CSA principal frame. $\mathrm{R_{ex}}$ is the additional relaxation rate caused by chemical exchange process, which only affects $\mathrm{R_2}$ since chemical exchange process usually happens on much slower timescale (~ms).

D.2 ^{15}N Relaxation Parameter Calculation from MD Trajectories

The calculation of backbone ^{15}N relaxation parameters from MD trajectories includes several following steps. The basic idea is to calculate the spectral density function $J(\omega)$ from MD trajectories first, then ^{15}N relaxation parameters can be obtained based on Equations (D.6)–(D.9).

(i) Since most MD simulations cannot predict molecular overall tumbling time τ_R correctly, the overall tumbling motion is usually removed from the MD trajectory by aligning all frames based on secondary structural regions (usually C^α atoms in those regions). However, when simulated molecules do not have well-folded structures (such as disordered proteins or peptides), this step should be skipped since it is difficult to define well-folded secondary structure regions for those molecules.

(ii) The vectors $\mathbf{r}$ formed by backbone amide group ^{1}H and ^{15}N are extracted per frame basis from the MD trajectory, the correlation function is calculated according to the following definition:

$$g(\tau) = \langle P_2[\hat{r}(t) \cdot \hat{r}(t+\tau)] \rangle = \frac{3\langle [\hat{r}(t) \cdot \hat{r}(t+\tau)]^2 \rangle - 1}{2} \tag{D.10}$$

$\hat{r}$ is the unit vector along the direction of N–H vector $\mathbf{r}$ ($\hat{r} = \mathbf{r}/|\mathbf{r}|$). The upper limit of τ is typically $\sim 80\%$ of total MD trajectory length to ensure enough sampling for calculating $g(\tau)$.

(iii) The correlation function $g(\tau)$ is fitted with six exponentials such that it has the form:

$$g(\tau) = \sum_{i=1}^{6} c_i \exp(-\tau/\tau_i) \quad \left(\sum_{i=1}^{6} c_i = 1 \right) \tag{D.11}$$

τ_i is the correlation time for the i-th component ($1 \leq i \leq 6$). Such decomposition makes subsequent calculation of spectral density function $J(\omega)$ much easier.

(iv) If overall tumbling motion was removed in the first step, it should be added back according to experimentally determined tumbling time τ_R at this step. After that the correlation function $C(\tau)$ becomes:

$$\begin{aligned} C(\tau) &= g(\tau)\exp(-\tau/\tau_R) = \sum_{i=1}^{6} c_i \exp(-\tau/\tau_i)\exp(-\tau/\tau_R) \\ &= \sum_{i=1}^{6} c_i \exp(-\tau/\tau_i^{\text{eff}}) \end{aligned} \tag{D.12}$$

τ_i^{eff} is the effective correlation time that already takes overall tumbling motion into account ($1/\tau_i^{\text{eff}} = 1/\tau_i + 1/\tau_R$). If step (i) was skipped then this step should be skipped as well.

(v) The spectral density function $J(\omega)$ is calculated from the correlation function $C(\tau)$ based on Equation (D.5). Since $C(\tau)$ is decomposed into six components, $J(\omega)$ can also be expressed as the sum of six components:

$$J(\omega) = \sum_{i=1}^{6} \frac{2\tau_i^{\text{eff}}}{5(1 + (\omega\tau_i^{\text{eff}})^2)} \tag{D.13}$$

(vi) Once the $J(\omega)$ has been determined, ^{15}N relaxation parameters can be calculated based on Equations (D.6)–(D.9).

All steps as described above have been followed to calculate ^{15}N relaxation parameters for Sos peptide within Sos:wtSH3 or Sos·dmSH3 complexes (including steps (i) and (iv)). For free Sos peptide, the two steps (i) and (iv) are skipped, instead another rescaling factor κ is introduced to correct motional timescale errors caused by TIP3P solvent[444]. After applying the rescaling factor κ it produces $\tau_i^{\text{eff}} = \kappa\tau_i$, therefore all motional components are slowed down as $\kappa > 1$. The rescaling factor κ is determined mainly based on the agreement between MD simulation results and experimental data, which is finally chosen as $\kappa = 2.68$ for the MD simulation of free Sos peptide in TIP3P solvent.

VITA

VITA

Tairan Yuwen was born in Beijing, China, on June 5, 1986. He also grew up in Beijing where he has received most of the education. He enrolled in Peking University in 2004 and received the Bachelor's Degree of Biological Sciences in 2008. He finished Bachelor's Degree thesis in Dr. Luhua Lai's lab in the Chemistry Department of Peking University about amyloid fibril inhibition with small heat-shock proteins.

After graduation from Peking University, he came to USA and entered PULSe program of Purdue University in August 2008. After one year's lab rotations he joined Nikolai Skrynnikov group and finished several research projects related to the topics of NMR pulse sequence development and MD simulation for biomolecules. He is currently a candidate for the Doctor of Philosophy Degree in the Department of Chemistry.

PUBLICATIONS

J Biomol NMR (2011) 51:131–150
DOI 10.1007/s10858-011-9548-7

ARTICLE

Domain cooperativity in multidomain proteins: what can we learn from molecular alignment in anisotropic media?

Tairan Yuwen · Carol Beth Post · Nikolai R. Skrynnikov

Received: 9 May 2011 / Accepted: 7 July 2011

Abstract Many proteins have modular design with multiple globular domains connected via flexible linkers. As a simple model of such system, we study a tandem construct consisting of two identical SH3 domains and a variable-length Gly/Ser linker. When the linker is short, this construct represents a dumbbell-shaped molecule with limited amount of domain–domain mobility. Due to its elongated shape, this molecule efficiently aligns in steric alignment media. As the length of the linker increases, the two domains become effectively uncoupled and begin to behave as independent entities. Consequently, their degree of alignment drops, approaching that found in the (near-spherical) isolated SH3 domains. To model the dependence of alignment parameters on the length of the interdomain linker, we have generated *in silico* a series of conformational ensembles representing SH3 tandems with different linker length. These ensembles were subsequently used as input for alignment prediction software PALES. The predicted alignment tensors were compared with the results of experimental measurements using a series of tandem-SH3 samples in PEG/hexanol alignment media. This comparison broadly confirmed the expected trends. At the same time, it has been found that the isolated SH3 domain aligns much stronger than expected. This finding can be attributed to complex morphology of the PEG/hexanol media and/or to weak site-specific interactions between the protein and the media. In the latter case, there are strong indications that electrostatic interactions may play a role. The fact that PEG/hexanol does not behave as a simple steric media should serve as a caution for studies that use PALES as a quantitative prediction tool (especially for disordered proteins). Further progress in this area depends on our ability to accurately model the anisotropic media and its site-specific interactions with protein molecules. Once this ability is improved, it should be possible to use the alignment parameters as a measure of domain–domain cooperativity, thus identifying the situations where two domains transiently interact with each other or become coupled through a partially structured linker.

Keywords Multidomain proteins · Residual dipolar couplings · Alignment tensor · Generalized degree of order · Domain-domain motion · Flexible linker · Conformational ensemble · Conformational disorder · Tandem SH3 construct · PALES software · Steric alignment · Electrostatic alignment · PEG/hexanol alignment media

Electronic supplementary material The online version of this article (doi:10.1007/s10858-011-9548-7) contains supplementary material, which is available to authorized users.

T. Yuwen · N. R. Skrynnikov (✉)
Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA
e-mail: nikolai@purdue.edu

C. B. Post
Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47907, USA

Introduction

Modular design is a hallmark of eukaryotic proteins. A great number of them contain adaptor domains, such as SH2, SH3, WW, PDZ, FHA, etc., connected through stretches of predominantly flexible linker. These domains present a number of 'generic' binding sites, facilitating assembly of protein complexes that prove to be of immense importance in signal transduction. Binding of adaptor domains to their molecular

targets is typically characterized by relatively low affinity (low μM) and limited specificity. As a consequence, the complexes often have the ability to rapidly dissociate, offering fast dynamic response to external stimuli. Because many domains can bind multiple targets, this mechanism gives rise to sophisticated, non-linear signaling networks, efficiently serving the needs of higher organisms (Cesareni et al. 2005).

As an illustration, let us briefly discuss the adaptor protein c-Crk. In mammalian cells, c-Crk is found in two different isoforms derived from alternative splicing. The smaller species, c-Crk I, are comprised of an SH2 domain followed by a ~15-residue linker and an SH3 domain (N-SH3). The larger form, c-Crk II, additionally contains a ~50-residue linker and an extra SH3 domain (C-SH3) (Kobashigawa et al. 2007).

Through its SH2 domain, Crk binds a variety of phosphotyrosine-containing proteins, most prominently p130Cas. A number of important signaling pathways converge at the p130Cas/Crk assembly, which contains multiple copies of Crk and thus has an ability to bring together a number of different Crk binding partners. The major role of these (interconnected) pathways in cell growth, motility, proliferation, adhesion, and apoptosis has been well documented (Feller 2001). There is also a strong link between Crk-mediated signaling processes and various human cancers, including lung and breast cancers, soft-tissue sarcomas and brain cancers (Linghu et al. 2006; Nishihara et al. 2002; Miller et al. 2003; Takino et al. 2003; Watanabe et al. 2006).

From a structural perspective, one of the key questions that can be asked about Crk and other modular proteins concerns the interactions between domains. On one hand, there is a scenario where two domains, connected through long and flexible linker, behave as essentially independent entities. The domain–domain coupling in this case is minimal. On the other hand, two domains can be significantly coupled through direct interaction and/or a conformationally restricted linker. In the case of Crk adaptor, recent structural study by Inagaki and co-workers showed that c-Crk I conforms more closely to the first scenario, while c-Crk II represents the second type of situation (Kobashigawa et al. 2007). More specifically, in the c-Crk I the two constitutive domains, SH2 and N-SH3, display significant motional freedom, whereas in the c-Crk II the three domains, SH2, N-SH3, and C-SH3, form a compact structure centered around the piece of the inter-SH3 linker. This result highlights the role of the extra domain: the binding surface in Crk C-SH3 has an unusual amino-acid composition, which degrades its ability to bind PxxP ligands; at the same time, C-SH3 occludes the binding site in N-SH3, thus downregulating a number of signaling events (Ogawa et al. 1994; Akakura et al. 2005; Muralidharan et al. 2006; Sarkar et al. 2007; Kobashigawa et al. 2007). Another domain rearrangement, which has the effect of shutting down the Crk-mediated signals, occurs upon phosphorylation of the tyrosine residue in the inter-SH3 linker. The subsequent (intramolecular) binding of the pYxxP motif to the SH2 domain abrogates most of the Crk interactions (Donaldson et al. 2002; Cipres et al. 2007; Peterson and Long 2008; Kobashigawa et al. 2007).

The above example underscores the importance of domain–domain coupling. Note, however, that it is not always possible to investigate domain interactions by straightforward structural methods. Indeed, as already mentioned, these interactions are often dynamic in nature, corresponding to diverse and short-lived conformational species. Two different approaches have been developed to address this situation. The first method relies on backbone ^{15}N relaxation. Indeed, domain cooperativity leads to increases in rotational correlation time τ_R, which can be sensitively detected by relaxation experiments (Zhang et al. 2008; Bae et al. 2009). This effect has been recently explored in-depth by Walsh and co-workers (Walsh et al. 2010) using protein constructs that are similar to those employed in our work. An accurate interpretation, however, can be challenging because spin relaxation is sensitive to many motional modes, including domain–domain motions (Chen and Tjandra 2008; Wong et al. 2009). It can also be difficult to 'calibrate' this method since the existing algorithms for prediction of τ_R have limited accuracy (de la Torre et al. 2000). The second method makes use of residual dipolar couplings. In this approach, one of the domains is loaded with a paramagnetic ion, thus achieving a weak alignment in the external magnetic field. If the two domains are fully coupled, then the second domain shows the same degree of alignment as the first; conversely, if the domains are uncoupled, then the second domain fails to show any alignment (Bertini et al. 2004; Bertini et al. 2007). While elegant and potentially useful, this method faces certain technical difficulties as it depends on suitably engineered metal-binding sites or carefully designed chelate tags (Ikegami et al. 2004; Rodriguez-Castaneda et al. 2006; Keizers et al. 2007).

In this study we turn to more conventional alignment methods involving steric mechanism. We further focus on the important limiting case where a pair of non-interacting domains is connected through a variable-length flexible linker. To model such a system, we manufactured a series of chimera proteins consisting of two copies of chicken α-spectrin SH3 domain (α-spc SH3) (Musacchio et al. 1992) connected through a flexible $(GGSGG)_n$-type linker (Huston et al. 1988). In the case of the short linker, this construct represents a dumbbell-shaped molecule which becomes efficiently oriented in a steric alignment media. Conversely, if the linker is long, then each SH3 domain behaves as an independent entity. Furthermore, since each

individual domain has a nearly spherical shape, it displays only a small degree of alignment. Thus, we expect that the degree of alignment changes in a sigmoidal fashion—from high values (short linker) to low values (long linker). The broad goal of this paper is to quantitate this dependence, both experimentally and by means of simulations. In this manner we seek to establish an important reference case—that of a protein with minimal domain–domain coupling (due to a trivial linker connectivity). Anything beyond this minimal coupling would point toward domain cooperativity. Such cooperativity can arise from domain–domain interactions, an increasingly structured linker region, or a combination of these two factors.

Materials and methods

Plasmid construction

The original pET3d plasmid encoding chicken α-spectrin SH3 domain (full-length construct fSH3) was a gift from Prof. Bernd Reif. Since the N-terminus of α-spc SH3 is highly disordered (Musacchio et al. 1992; Blanco et al. 1997; Chevelkov et al. 2005), we truncated the construct by deleting residues 2–6 (deletion construct dSH3). We have also manufactured three chimera proteins, each of which contained two copies of dSH3 connected via a glycine-serine linker. To link the domains we have used the GSG, GG(GGSGG)GG, and GG$(GGSGG)_5$GG sequences (the latter contains 29 residues). In what follows, these tandem constructs are referred to as dSH3-*sl*-dSH3, dSH3-*ml*-dSH3, and dSH3-*ll*-dSH3, alluding to *s*hort, *m*edium-length, and *l*ong *l*inker, respectively. To produce the cDNA for tandem constructs, two separate PCR reactions were carried out. One PCR product encoded N-terminal dSH3 plus half of the linker region; it was digested with restriction enzymes *NcoI* and *BspEI*. The other PCR product encoded the remaining half of the linker plus C-terminal SH3 domain; it was digested with *BspEI* and *BamHI*. In the next step, the two digested PCR products were cloned into pET3d vector and, at the same time, fused by means of the T4 DNA ligase. Note that the use of *BspEI* sites and T4 DNA ligase does not introduce any unwanted residues in the linker sequence. All plasmids were transformed into BL21(DE3) *E. coli* strain and sequenced to check for accuracy. Primer sequences used to produce the DNA vectors and amino-acid sequences of the SH3 constructs are listed in Table S1.

Protein expression and purification

The bacteria transfected with plasmids were grown in 1 L of M9 minimal media enriched with $^{15}NH_4Cl$ at the temperature 37°C. Protein expression was induced at $OD_{600} \sim 0.6$ by adding 0.5 mM IPTG. After 4 h of incubation the cells were harvested by centrifugation (15 min at 8,000 rpm) and stored at −80°C. The cell pellet was then resuspended in pH 8 lysis buffer and lysed using a French Press. fSH3 was contained in the supernatant; it has been purified using the previously described protocol (Chevelkov et al. 2005). All other constructs formed inclusion bodies during the expression. To purify the precipitated protein material, we first washed the French press lysis pellet four times using 20 mM Tris (pH 8), 1 mM EDTA, 100 mM NaCl, 0.5% Triton X-100 buffer. Each wash was carried out in 30 mL of the buffer (30 min at 18,000 rpm); the last wash was done without the Triton. The pellet was then dissolved in 30 mL of 8 M urea, 20 mM Na-citrate, pH 3.5 denaturing buffer and heated in a water bath at 70°C for 30 min. The solution was subsequently dialyzed over the course of 1 day against 20 mM Na-citrate, pH 3.5 refolding buffer using the dialysis bag with MW cutoff 3 kDa. The refolded protein was purified using the Superdex 75 gel filtration column and then concentrated using the Millipore Amicon devices with 3 kDa MW cutoff. All samples have been analyzed by SDS–PAGE to check the protein molecular weight. Protein concentration determined by UV absorbance at 280 nm was in the range from 0.5 to 1.0 mM. In addition to α-spectrin SH3 domain, we have also prepared two samples of human ubiquitin. The plasmid was supplied by Rachel Klevit laboratory through Addgene repository; the expression and purification procedure was adapted from (Lazar et al. 1997).

NMR Spectroscopy

The experimental measurements involved five constructs of α-spectrin SH3 domain (fSH3, dSH3, dSH3-*sl*-dSH3, dSH3-*ml*-dSH3, and dSH3-*ll*-dSH3) in solution with 20 mM Na-citrate, pH 3.5, 0.02% (w/v) NaN_3. One extra sample of fSH3 additionally included 100 mM NaCl (see Table 1). The oriented samples were prepared by using 5% penta(ethylene glycol) monododecyl ether (C12E5) and 1-hexanol (Rückert and Otting 2000). Two series of samples were prepared with the C12E5-to-hexanol molar ratio $r = 0.85$ and 0.96, respectively. The former condition offers stable alignment around the working temperature 25°C (Rückert and Otting 2000); the latter is the most popular choice in the studies employing PEG-hexanol media (Ulmer et al. 2003; Bewley 2001; Braddock et al. 2002). Special care was taken to maintain the same alignment conditions throughout each series of samples. Toward this goal, we used the stock solutions containing 10% C12E5 and $r = 0.85$ (0.96) fraction of 1-hexanol. The anisotropic samples were prepared by mixing 150 μL aliquot of the stock solution with 150 μL aliquot of the protein solution (to be placed in a Shigemi tube). Each freshly prepared batch of protein was used to make two or three NMR samples, including some replicate

Table 1 Alignment parameters from a series of samples in $r = 0.85$ PEG/hexanol media

	dSH3[a]	dSH3	fSH3	fSH3 (high salt)[b]	dSH3-*s*/-dSH3	dSH3-*s*/-dSH3[c]	dSH3-*m*/-dSH3	dSH3-*l*/-dSH3[a]	dSH3-*l*/-dSH3[c]
A_a (10^{-3})	2.23	3.01	-1.40	-2.27	-1.56	-1.51	0.67	0.65	0.75
R	0.62	0.40	0.37	0.39	0.37	0.36	0.65	0.27	0.29
$\{\alpha,\beta,\gamma\}$[d]	308° 62° 81°	308° 65° 80°	32° 38° 320°	27° 40° 325°	33° 80° 63°	34° 80° 63°	168° 53° 159°	107° 54° 155°	106° 54° 155°
N	29	25	46	27	42	42	39	45	45
Q	0.25	0.30	0.22	0.21	0.22	0.21	0.27	0.26	0.27
$\Delta(^2H)$ (Hz)	24.9	21.8	24.7	25.8	23.3	23.1	16.9	23.5	24.4
GDO (10^{-3})[e]	2.03[f]	2.93	1.19[f]	1.85	1.41[f]	1.37	0.92[f]	0.57[f]	0.63
NSP[g]	1.0	0.98	0.94	0.94	-0.53	-0.53	-0.32	0.26	0.27

The results from replicate samples are shown in shaded columns

[a] The RDC fits are shown in Fig. 6

[b] 100 mM NaCl added to the buffer solution

[c] The results included in Fig. 7a, b (green triangles)

[d] Relative to the crystallographic coordinates 1U06. The rotation as defined in PALES is inverse of the rotation used by other programs (Skrynnikov et al. 2000; Valafar and Prestegard 2004)

[e] Normalized toward $\Delta(^2H) = 20$ Hz

[f] Values used to optimize the electrostatic alignment model

[g] Relative to dSH3

samples. Every sample was allowed to equilibrate in the magnet for 2 h, before the residual quadrupolar coupling from 10% D_2O/90% H_2O solvent, $\Delta(^2H)$, was carefully measured. The $^1H^N$-^{15}N residual dipolar coupling data were collected using the IPAP-HSQC pulse sequence (Yang and Nagayama 1996; Ottiger et al. 1998) on a Varian Inova 600 spectrometer equipped with a triple-resonance probe and z-axis gradient. Data sets for isotropic and partially aligned samples were collected as (192, 576) complex matrices with spectral widths of 1,920 and 9,000 Hz in nitrogen and proton dimensions, respectively. Each experiment took approximately 2 h. After the data were collected, the residual ^2H coupling was re-measured; the $\Delta(^2H)$ drift never exceeded 0.2 Hz. The IPAP spectra were processed using the NMRPipe software package and peak positions were determined by means of the nlinLS fitting routine (Delaglio et al. 1995). In the case of tandem constructs, most of the peaks from the N-terminal domain are neatly overlapped with their counterpart peaks from the C-terminal domain. Hence, the RDC values measured for these residues represent the average of the N- and C-domain couplings. For a small number of sites that give rise to two resolved peaks, we choose to average the experimental RDCs on a pairwise basis (see Experimental Results). The experimental data from the tandem constructs represent, therefore, the average values (listed in Tables S2, S3). In analyzing the results, all datasets have been restricted to the core portion of the SH3 domain, residues 9–60.

Structural models

Each individual α-spectrin SH3 domain was modeled based on the crystallographic structure 1U06 (resolution 1.49 Å) (Chevelkov et al. 2005). Protons and several missing heavy atoms from the side chain of residue D48 were added to the structure and optimized using the facilities of CHARMM (Brooks et al. 2009) and XPLOR-NIH (Schwieters et al. 2003) with the CHARMM22 force field. The resulting domain coordinates were used to build structures of tandem dSH3 and other constructs. The structure 1U06 contains five side chains that are modeled with two alternate conformations. In our main structural model all of them are placed in the conformation found in the chain A of the crystallographic

structure. We have also tested other models, in particular the one where the ionizable E17 side chain is placed in conformation B. The energy of this alternative structure, as evaluated in implicit generalized Born solvent (Dominy and Brooks 1999), was found to be within 6 kcal/mol from the all-A structure. The structure, labeled BAAAA, was therefore deemed suitable for the analyses. In addition, the NMR structure 1AEY (Blanco et al. 1997) was also used.

As an example of the SH3 tandem, let us consider the structural model for dSH3-*ml*-dSH3. The amino acid sequence for this construct is M1(E7…L61)D62-GGGG SGGGG-(E7…L61)D62. The bracketed portions of the sequence correspond to the structured portion of the domains, as represented by the crystallographic coordinates 1U06. The remaining part, comprising the 10-residue linker and single-residue termini, is modeled as a random coil. To generate the coordinates of these segments, we used the program TraDES (routine foldtraj, default input parameters) (Feldman and Hogue 2000). For instance, in order to add N-terminal residue M1 we first generate a random dipeptide with the sequence ME. This peptide is then 'glued' onto the structure of the SH3 domain by overlaying the E residue of the peptide with E7 residue in the SH3 domain (specifically, N, C^{α}, and C' atoms are superimposed in this particular order). Finally, all redundant atoms are deleted. This includes all atoms from the peptide residue E, with the exception of H^{N}. The equivalent procedure is used for the C-terminal residue D62. In this case, we begin with the random dipeptide LD, which is 'glued' to the SH3 domain by overlaying the L residue of the peptide with L61 from the SH3 domain (C', C^{α}, and O atoms are superimposed). Subsequently, all atoms from the placeholder peptide residue L are deleted. The same prescriptions are applied to insert the ten-residue linker segment, DGGGGSGGGG.

The dSH3-*ml*-dSH3 model obtained in this fashion is subjected to a further round of MD refinement (1 ps at 100 K). The goal of the refinement is to relieve some of the strain caused by the 'structure assembly' procedure. The relative position of the domains is preserved during this stage (for this purpose all atoms within the domains have been assigned an artificially large mass, 10 kDa). The refinement protocol is implemented using the torsional angle dynamics option in XPLOR NIH (IVM module (Schwieters and Clore 2001)); the internal coordinates of the domains (all dihedral angles in the residues from 7 to 61, except for the angle φ in residue E7) are fixed and the planarity of peptide planes is enforced. This algorithm is used to produce the ensemble of 10,000 models for dSH3-*ml*-dSH3. Many of these models display severe steric clashes that cannot be resolved by our limited refinement procedure. To weed out these structures, we evaluate the energies of each model and analyze the resulting energy distribution as discussed below.

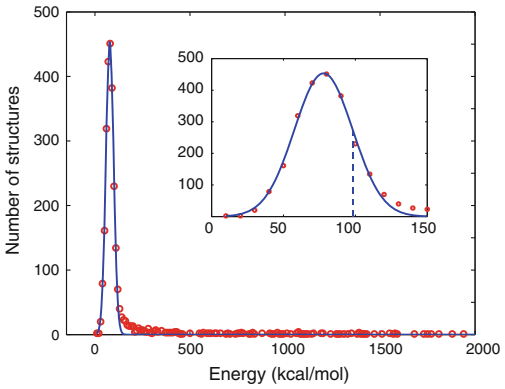


Fig. 1 Part of the energy distribution for 10,000-strong ensemble of dSH3-*ml*-dSH3 models (*red circles*) fitted with the Gaussian function (*blue curve*). The inset displays the expanded portion of the main graph; the *dashed line* corresponds to one standard deviation σ of the Gaussian curve

Figure 1 shows a portion of the energy distribution for 10,000-strong ensemble of dSH3-*ml*-dSH3 models. The energies are evaluated using XPLOR-NIH/CHARMM22 in vacuo (the solvation term is inconsequential, as we seek to identify the structures with severe steric clashes). The low-energy part of the distribution follows the Gaussian curve (see inset in Fig. 1). This is to be expected, since the distribution arises from variability in the random-coil linker region. The Gaussian energy distribution for disordered protein systems can be predicted from general considerations and also has been derived from a number of theoretical models (Bryngelson and Wolynes 1987; Shakhnovich and Gutin 1989; Bryngelson et al. 1995).

In Fig. 1, those structures that comply with the Gaussian distribution constitute a *bona fide* conformational ensemble, whereas the high-energy structures outside the Gaussian distribution (appearing to the right in the plot) suffer from steric conflicts. To address this problem, we analyzed the part of the energy distribution to the right of the dashed line (corresponding to one standard deviation σ of the fitted Gaussian curve) and eliminated a fraction of structures that fall outside the curve. In this manner the original ensemble of 10,000 models was trimmed down to 2,371 models. The retained structures are (1) free of significant steric clashes and (2) have their two domains positioned according to the conformation of the (random-coil) linker, as generated by TraDES. Several randomly selected structures from the resulting reduced ensemble are shown in Fig. 2.

The same procedure as described above has been used to build the models for other chimera proteins, as well as fSH3 and dSH3 (the former carries a 6-residue unstructured

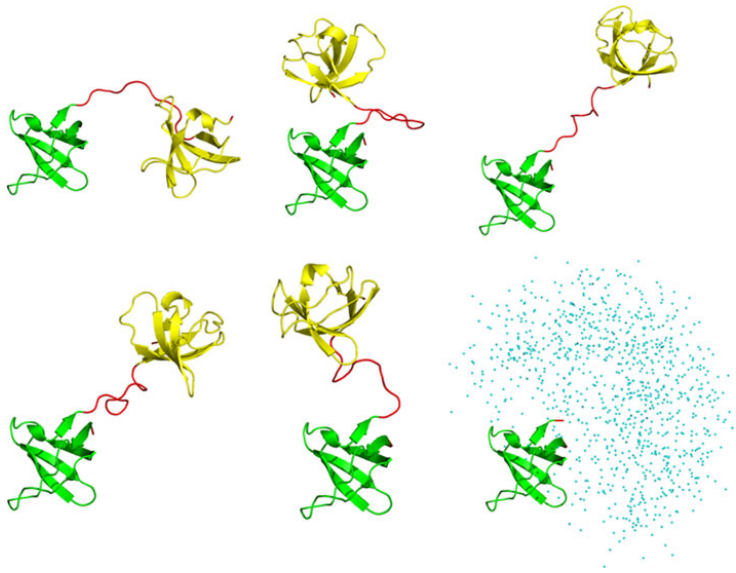


Fig. 2 Five representative structures from the final dSH3-*ml*-dSH3 ensemble. The N- and C-terminal SH3 domains (backbone coordinates same as in the crystallographic structure 1U06, residues 7–61) are plotted in green and gold, respectively; the regions modeled as random coil are colored in *red*. The panel in the lower right corner shows the spatial distribution of the centers of mass of the C-terminal domain in the coordinate frame of the N-terminal domain (as plotted for the entire ensemble)

N-terminus and a single-residue C-terminus, the latter contains two single-residue termini).

Alignment parameters

Protein alignment in liquid–crystalline media is traditionally characterized by the degree of alignment A_a, rhombicity R, and three Euler angles that specify the orientation of the principal axes of the alignment tensor relative to the molecular frame (Tjandra and Bax 1997). These parameters, however, are not necessarily best suited for drawing a comparison between different aligned samples (which becomes especially obvious when Euler angles are compared). For a suitable alternative we turn to the five-dimensional vector space spanned by the irreducible components of the Saupe alignment tensor (Moltke and Grzesiek 1999; Sass et al. 1999; Prestegard et al. 2000). In this linear space the scalar product between the two vectors formed from the Saupe matrix elements can be defined as:

$$\left\langle \mathbf{S}^{sample1} \middle| \mathbf{S}^{sample2} \right\rangle = \sum_{\substack{i=x,y,z \\ j=x,y,z}} S_{ij}^{sample1} S_{ij}^{sample2} \tag{1}$$

where S_{ij} are the elements of the 3×3 Saupe matrices associated with the rigid molecule or molecular fragment (e.g. SH3 domain within the tandem construct). *sample1* and *sample2* designations refer to samples involving different alignment media or different protein constructs, that share the same fragment (SH3 domain). Note that both $\mathbf{S}^{sample1}$ and $\mathbf{S}^{sample2}$ should be expressed in the same coordinate frame. In our case this frame is linked to the uniquely defined coordinates of the SH3 domain (pdb frame of the structure 1U06).

Using the definition in (1), one can introduce the Generalized Degree of Order (GDO) (Tolman et al. 2001) that characterizes the magnitude of alignment for a given sample:

$$\mathrm{GDO} = \sqrt{(2/3)\left\langle \mathbf{S}^{sample1} \middle| \mathbf{S}^{sample1} \right\rangle} = |A_a| \sqrt{1 + (3/4)R^2}. \tag{2}$$

Furthermore, if the scalar product (1) is normalized, then it can be viewed as a generalized measure for relative orientation of the two alignment tensors:

$$\mathrm{NSP} = \frac{\left\langle \mathbf{S}^{sample1} \middle| \mathbf{S}^{sample2} \right\rangle}{\sqrt{\left\langle \mathbf{S}^{sample1} \middle| \mathbf{S}^{sample1} \right\rangle \left\langle \mathbf{S}^{sample2} \middle| \mathbf{S}^{sample2} \right\rangle}}. \tag{3}$$

The values of the normalized scalar product (NSP) close to 1.0 indicate that the two alignment tensors differ only by a scaling factor, whereas the values around 0.0 suggest that the alignment frames are 'orthogonal'. The situation where $\mathrm{NSP} = -1$ can be described as 'antiparallel' alignment—the rhombicities and alignment frames are in this case identical, while the degrees of alignment A_a have opposite signs. This situation is observed in stretched versus compressed gels (Sass et al. 2000; Mohana-Borges et al. 2004), conventional versus 'flipped' bicelles (Opella and De Angelis 2007; Koenig et al. 1999), and other similar systems (Denisov et al. 2010). Although NSP defined by

(3) depends also on rhombicity R, it is mainly a measure of alignment axes orientation.

The definition of the Saupe tensor used here (Sass et al. 1999) differs by a factor of 2 from the one used elsewhere (Bax et al. 2001). The definitions of the GDO and NSP, however, are unambiguous. The numeric results for Saupe matrices, as required for evaluating Eqs. (2) and (3), are reported by the program PALES (Zweckstetter and Bax 2000; Zweckstetter 2008), both in the case of the experimentally determined (fitted) and predicted alignment tensors. Finally, the quality factor is defined as (Cornilescu et al. 1998; Bax 2003):

$$Q = \frac{\left(\sum_{i=1}^{N} (D_i^{exptl} - D_i^{calc})^2/N\right)^{1/2}}{\left(\sum_{i=1}^{N} (D_i^{exptl})^2/N\right)^{1/2}}. \quad (4)$$

PALES predictions

The alignment mechanism in PEG/hexanol media is commonly assumed to be steric (Rückert and Otting 2000; Zhang and Zuiderweg 2004; Fischer and Geyer 2005; Maltsev et al. 2008; Berlin et al. 2009). To model this media we used the PALES procedure originally developed for DMPC/DHPC bicelles (Bax and Tjandra 1997). The thickness of the planar bilayer was set to 28.6 Å (Freyssingeas et al. 1996). It is worth mentioning that bicelle calculations in PALES have been programmed for DMPC/DHPC, so that the code implicitly accounts for the presence of 5 mM free DHPC in the solvent (Ottiger and Bax 1998; Simon et al. 2005). To correct for this small contribution and, in addition, account for 0.3 wt% of free hexanol in solution (Freyssingeas et al. 1996) one should use a slightly altered value of the liquid crystal concentration. For instance, in the PALES calculations aimed at 5% PEG, $r = 0.85$ media the effective liquid crystal concentration should be set to 65 mg/ml. The order parameter of the PEG/hexanol bicelle was assumed to be the same (0.8) as for DMPC/DHPC.

We have also attempted to predict the alignment parameters under the assumption that an electrostatic mechanism plays a significant role. The corresponding PALES calculations involve several additional variables: charge density on the surface of the bilayer σ (treated as adjustable parameter), sample pH (actual value 3.5; also treated as adjustable parameter), and ionic strength of the solution (17 mM, corresponding to 20 mM citrate buffer at pH 3.5). To determine the charges on α-spc SH3, we have used the program PROPKA, which calculates the pK_a values of ionizable groups (Li et al. 2005; Bas et al. 2008). Since there is significant structural variability in tandem constructs and fSH3, we have run PROPKA on each individual conformer from the structural ensemble. The results (i.e. the coordinates of a given conformer plus the charges calculated for this particular conformer) were then used as input for PALES calculations. In principle, more sophisticated methods for calculating charges can be used (Gordon et al. 2005; Khandogin and Brooks 2006) and ultimately the pK_a values can be determined experimentally (Tollinger et al. 2002; Andre et al. 2007). However, such effort seems unwarranted given that the models used for prediction of electrostatic alignment remain relatively crude (Zweckstetter et al. 2004; Zweckstetter 2006; Wu et al. 2006). Of note, we have found a programming error in the PALES module responsible for computations of electrostatic alignment. Specifically, natural logarithm is mistakenly used in the Henderson-Hasselbalch equation instead of the base 10 logarithm. As a result, given the set of pK_a (either default or user-supplied values) the program fails to correctly calculate point charges at a given pH. The problem can be circumvented by using a table of user-supplied point charges as the PALES input. In the latest revision of PALES this issue has been corrected (M. Zweckstetter, personal communication, 2011).

Dipolar couplings predicted for the individual conformers are averaged, resulting in a single set of simulated RDCs. When drawing comparison with the experimental data from the tandem constructs, we further average the computed couplings from the N- and C-terminal domains (in accordance with our experimental protocol, see above). Note that such averaging procedure is fully justified so long as both domains have the same (invariant) internal structure. Indeed, RDCs by definition represent the average between many molecular orientations. The resulting dataset is subsequently restricted to the core portion of the α-spc SH3 domain, residues 9–60, and fitted with the (protonated) structure 1U06. In the situation when residues 9–60 are represented by one and the same set of coordinates throughout the course of the calculations, the residual χ^2 from such fitting procedure is strictly zero. The alignment parameters extracted in this fashion are then used to calculate the theoretical values of the GDO and NSP, Eqs. (2) and (3). Note that the averaging of simulated RDCs with respect to the N- and C-terminal domains affects the final GDO and NSP values.

Theory: general outline

Let us consider a generalized two-domain protein, where a pair of globular, near-spherical domains is connected via a flexible structureless linker. In the case when the linker is short, i.e. consists of no more than several residues, the protein has a dumbbell shape. Furthermore, a short linker allows for little motional freedom so that domain–domain motion is highly restricted. To a good approximation, the

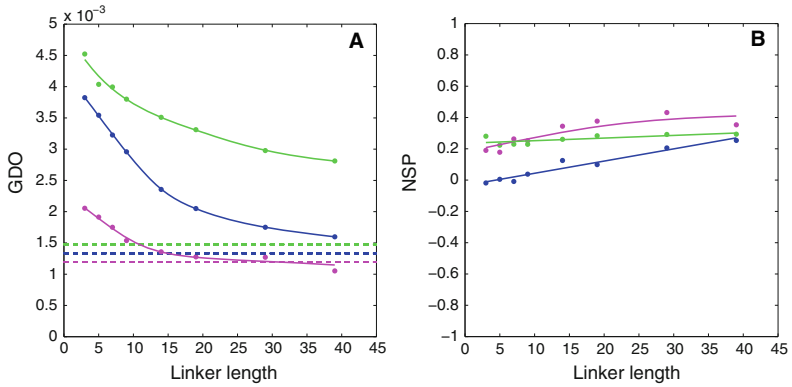


Fig. 3 PALES simulations of steric alignment in a series of computer-generated tandem proteins. To generate structural ensembles, the two identical domains (represented by crystallographic or NMR coordinates and presumed to be fully rigid) were directly connected via glycine/serine linkers. The simulated couplings from all non-proline residues in the N-terminal domain were used to derive the alignment parameters. Other details of the structure generation and RDC simulations are as described in the Materials and Methods. Shown are the results for the tandem constructs of the disintegrin domain (structure 1MPZ (Moreno-Murciano et al. 2003), 41 residue, radius of gyration 8.8 Å, *magenta symbols*), SH2 domain (extracted from structure 1A81 (Futterer et al. 1998), 106 residues, 13.0 Å, *blue*), and nucleotide-binding domain (3GWI (Hakansson 2009), 170 residues, 15.3 Å, *green*). **a** Simulated GDO values for the domains which belong to the tandems (*circles*) and the corresponding isolated domains (*dashed lines*). **b** Simulated NSP values characterizing the mutual projection of the alignment tensors from (1) the domain which belongs to the tandem and (2) the corresponding isolated domain. The spline curves are added to the graphs for visual guidance. To verify the convergence of the simulations, we have regenerated this plot using a different set of randomly generated structural ensembles. The results are shown in Fig. S1

protein can be described as a rigid body with highly elongated shape. Being placed in a steric alignment media, this protein should efficiently align and display a high GDO value, (2). The long axis of the alignment tensor should roughly coincide with the line connecting the centers of mass of the two domains. Generally speaking, the alignment PAS in such a two-domain construct has no relation to the PAS of the individual (near-spherical) domain. Therefore the NSP between the two respective tensors (3), should deviate from 1.0 (unless by coincidence).

Let us now consider the opposite situation, where the flexible linker is very long. In this case the two domains move almost freely and no longer sense each other's presence. Their GDOs, therefore, should approach that of an individual domain. Given that the shape of each domain is nearly spherical, the GDO should drop to a low value. Likewise, the PAS orientation should generally approach that found in the individual domain. Consequently, the scalar product between the two alignment tensors—the one from the domain which belongs to the tandem and the one from the isolated domain—should converge toward 1.0.

Note that neither GDO, nor NSP are expected to show a quantitative convergence toward the single-domain limit. Indeed, the behavior of a tethered domain is similar, but not identical, to that of the isolated domain. In particular, if the shape of the domain is very close to spherical then even a small alteration of the shape (such as caused by attachment of the linker) can lead to a significant change in magnitude and especially orientation of the alignment tensor. Therefore, the convergence toward the single-domain limit should be rather viewed as a broad trend.

To illustrate this expected behavior, we have undertaken a series of PALES simulations for a number of protein domains, Fig. 3. For these simulations we have chosen the domains of different size with close-to-spherical shape; if desired, the results can be extended to domains with elongated or flattened shapes. The coordinates of the domains have been used to generate tandem structures, which were subsequently processed by PALES. As expected, increasing the linker length from three residues to five, then to seven, etc. leads to pronounced decreases in the GDO values. Eventually, each curve approaches a plateau which is close to, but distinct from the GDO of the isolated domain (dashed horizontal lines in Fig. 3a). Similarly, as the linker length is increased, the NSP value becomes closer to 1.0 (Fig. 3b). The trend, however, is weak, which means that the orientation of the alignment frame in the tethered domain remains quite different from the isolated domain.

It is interesting to discuss how the results scale with the size of the system. Generally, one would expect that for the smallest unit, disintegrin, the two domains become 'uncoupled' and the GDO (NSP) plateau is reached already with a moderately long linker. Conversely, for the biggest unit, nucleotide-binding domain (NBD), this does not happen until the linker becomes much longer. A rough theoretical estimate can be readily obtained for the scaling

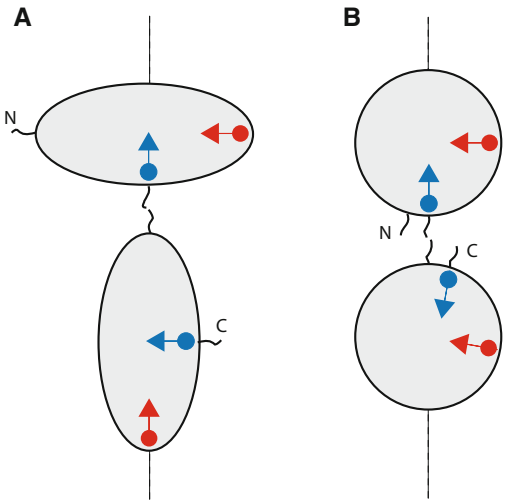


Fig. 4 Schematic structure of tandem domains with resident spins. *Blue circles* represent ^{15}N spins belonging to *i*-th residue in the N- and C-terminal domains (assuming that residue numbering is the same in the two domains); *blue arrows* represent the corresponding ^{15}N-^{1}H^{N} vectors. *Red circles/arrows* refer to the *j*-th residue. The *dashed line* corresponds to the long axis of the steric alignment tensor

properties of the two-domain constructs. The radii of gyration of the disintegrin and NBD domains are 8.8 and 15.3 Å, respectively. Let us accept that both domains are spherical. The average end-to-end distance in the random peptide chain, such as the computer-generated linker, scales with the number of residues as $N^{0.6}$ (Fitzkee and Rose 2004). From here we can estimate the 'scaling factor' that relates the NBD tandem to disintegrin tandem. For example, the NBD tandem with 25-residue linker can be viewed as a 'resized' copy of the disintegrin tandem with 10-residue linker. This is roughly consistent with the trends seen in Fig. 3a.

Generally speaking, N- and C-terminal domains in the tandem are inequivalent: one is attached to the linker through its C-terminal residues while the other through the N-terminal residues. In principle, this should lead to doubling of the spectral resonances. This situation is illustrated in Fig. 4a, which shows two domains connected in tandem. The scheme Fig. 4a suggests that some of the amide signals from N- and C-domains should be resolved with respect to their chemical shift. Furthermore, most of the signals should be distinguishable with respect to dipolar couplings. For instance, the equivalent residues from N- and C-terminal domains (blue circles) sense somewhat different chemical environment: one is located next to the fused terminus, while the other is close to the free terminus. Hence one can expect to observe two distinct HSQC peaks from these sites. Moreover, the orientation of the ^{15}N-^{1}H^{N} vectors at the two sites (blue arrows) relative to the long axis of the alignment tensor (dashed line) is also different. Hence the residual dipolar couplings for these two sites should be different as well. In fact, it is anticipated that most of the N- and C-terminal residues can be differentiated on the basis of their dipolar couplings. For instance, another pair of equivalent residues (red circles) illustrates the case where chemical shifts are likely to be degenerate, but RDCs are very different. Figure 4b represents a 'softer' version of the same scenario. When N- and C-termini of each domain are positioned close to each other, the RDCs from the two domains may turn out to be similar. Apparently, this scenario holds true for the α-spc SH3 tandem (see Experimental Results). Similar results were recently reported in the experimental study of GB1 construct featuring a short 3-residue linker (Walsh et al. 2010). Of the simulated constructs, see Fig. 3, NBD tandem illustrates the situation where the couplings from the two domains are reasonably close. The other two constructs, however, predict significant differences between the two domains.

Results

RDC data

The spectra of the dSH3 tandems turn out to be very similar to the spectrum of the isolated dSH3. For example, Fig. 5 shows the spectrum of the two-domain construct with long disordered linker superimposed on the spectrum of the single domain (spectral peaks colored red and blue, respectively). The resonances from multiple glycine residues of the random-coil GG(GGSGG)$_5$GG sequence are all grouped in the area of the spectrum where they do not interfere with intra-domain peaks (outlined by the green contour). Most of the peaks from N- and C-terminal domains overlap. Furthermore, they overlap or fall very close to the peaks from the isolated dSH3 domain (cf. red and blue contours in Fig. 5), so that the assignments (Blanco et al. 1997) can be transferred in a straightforward fashion. The transfer of resonance assignment from a single domain to a multidomain assembly is presently a common practice (Tzakos et al. 2006; Gelis et al. 2007; Sprangers and Kay 2007; Heikkinen et al. 2009; Brosey et al. 2009). The lack of resonance shifts between the isolated domain and the tandem confirms the absence of specific domain–domain interactions (along the same lines, ^{15}N relaxation study has not found any evidence of protein aggregation in α-spc SH3 (Chevelkov et al. 2007)).

As discussed above, a small number of residues that are located close to the linker attachment point give rise to a pair of resolved peaks, corresponding to the sites in the

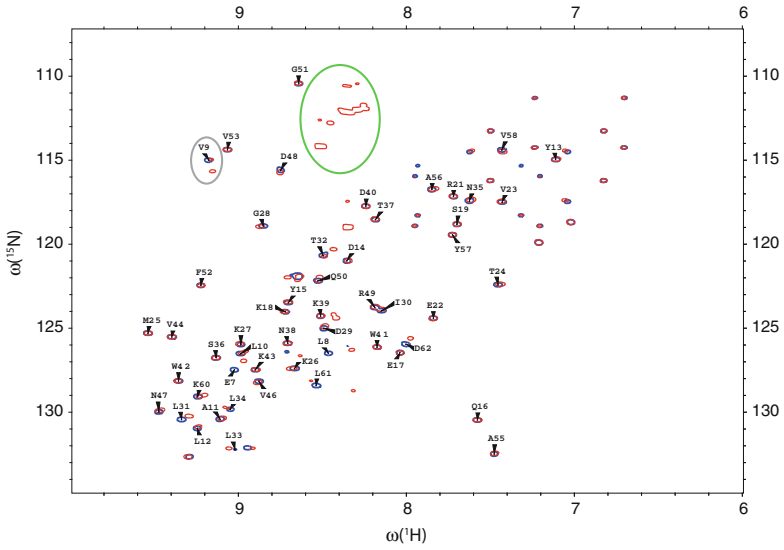


Fig. 5 HSQC spectra of the isolated dSH3 domain (*blue*) and dSH3-*ll*-dSH3 tandem (*red*). The dSH3 resonances are labeled except for Trp, Asn, and Gln side-chain correlations. The elliptical contours outline the resonances from the glycine-rich linker and from the residue V9 which demonstrates peak doubling effect

N- and C-terminal domains (one example of this behavior is residue V9, outlined by the grey contour in the spectral map Fig. 5). The effect is most pronounced in the short-linker construct, dSH3-*sl*-dSH3, where the two domains are spatially close. In this sample, 18 residues show the peak doubling effect, with half of them fully resolved. In particular, the peaks from terminal residues 7, 8 and 61, 62 are strongly affected. Inspection of the crystallographic structure shows that these two segments are connected: there is a backbone–backbone hydrogen bond between residues 8 and 61. Thus all of the four residues prove to be in the immediate vicinity of the linker region, resembling the arrangement shown in Fig. 4b. The split peaks associated with residues 7, 8, and 61, 62 show large chemical-shift separations, which makes it impossible to assign most of them. Furthermore, it is likely that the conformation of these residues in dSH3-*sl*-dSH3 differs from that seen in the crystallographic structure of fSH3. Considering these complications we choose to restrict the data analysis to the core region of the domains, residues 9–60. This restriction has been applied to all α-spc SH3 constructs investigated experimentally and via simulations.

A separate question arises as to whether the resonances from N- and C-terminal domains can be distinguished on the basis of residual dipolar couplings. In principle, one may expect that a pair of peaks that is completely overlapped in HSQC spectrum would become resolved in IPAP-HSQC (due to potentially large differences in RDCs, on the order of tens of Hz; see Fig. 4a and surrounding discussion). In reality, however, we have not found any examples of such behavior, despite high resolution of the IPAP-HSQC spectra. In the experiments involving dSH3-*sl*-dSH3, a few RDCs that have been measured separately for the sites in the N- and C-domains differ by no more than several Hz (e.g. in the case of V9 the difference is 0.5 Hz). At this level, RDCs alone cannot produce resolved spectral peaks. In summary, it appears that tandem dSH3 constructs behave as indicated in the scheme Fig. 4b: the N- and C-termini of each domain are aligned and together form a linker attachment point; the RDCs in the two domains are similar.

Given that only few residues allow for separate measurements of RDCs associated with N- and C-terminal domains and the measured values prove to be very similar, we choose to average such couplings and add the results to the bigger dataset consisting of the average RDCs. For instance, in the case of dSH3-*ll*-dSH3 we have measured 44 couplings that inherently represent the average values, plus two distinct couplings from the residue V9. After these two couplings are averaged, the final data set consists of 45 entries (pertaining to the region 9–60 in the dSH3 domain). The same approach has been applied to the other samples (Tables S2, S3).

The RDC data obtained as described above have been fitted to the crystallographic coordinates 1U06, chain A (see Materials and Methods). Although the experimental RDCs from the tandem constructs represent the averages between the N- and C-terminal domains, they have been fitted in the standard fashion, same as the data from the isolated dSH3 and fSH3 domains. As already commented, this approach is fully legitimate, so long as the internal structure of the two domains is identical. Indeed, residual

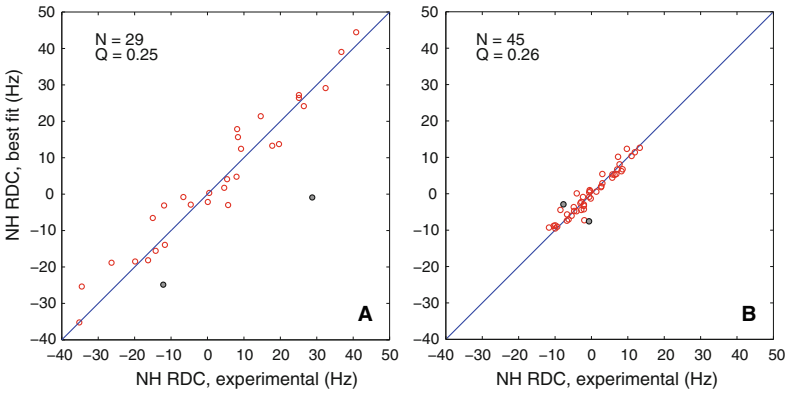


Fig. 6 Fits of the experimental RDCs from **a** the isolated dSH3 domain and **b** dSH3-*ll*-dSH3 tandem with the structural model based on the coordinate set 1U06 (protonated, energy minimized). The fitting is limited to the core portion of the protein, residues 9–60, and does not include the data from N47 and D48 (*grey circles* in the plot). The total number of the fitted RDCs and the figure of merit Q are indicated in the graphs

Table 2 Alignment parameters from a series of samples in $r = 0.96$ PEG/hexanol media

	dSH3	fSH3	dSH3-*sl*-dSH3	dSH3-*ml*-dSH3	dSH3-*ll*-dSH3
A_a (10^{-3})	2.23	−0.78	−1.50	−0.78	0.41
R	0.40	0.33	0.36	0.64	0.41
$\{\alpha, \beta, \gamma\}$	306°	32°	33°	36°	103°
	65°	39°	80°	83°	54°
	79°	321°	63°	64°	154°
N	43	48	41	40	45
Q	0.25	0.21	0.22	0.26	0.25
$\Delta(^2\text{H})$ (Hz)	17.1	12.3	20.3	15.0	11.0
GDO (10^{-3})	2.76	1.32	1.54	1.18	0.80
NSP	1.0	0.90	−0.64	−0.47	0.21

The conventions are the same as in Table 1

dipolar couplings represent, by definition, the average between many molecular orientations. The quality of the RDC fits for dSH3 and dSH3-*ll*-dSH3 is illustrated in Fig. 6a and b, respectively. In both cases the agreement between the experimental and fitted couplings is good, as evidenced by the Q factors of 0.25 and 0.26 (typical of a small globular protein represented by a crystallographic structure with 1.5 Å resolution) (Bax 2003).

Two residues consistently show poor agreement in all fits—N47 and D48 (grey circles in Fig. 6). This discrepancy should be attributed to the local conformational variability at this particular site. In the crystal, N47 assumes an unusual conformation ($\phi \approx 50°, \psi \approx -110°$) (Vega et al. 2000) which has a significant destabilizing effect on the structure (Martinez et al. 1998). A similarly unfavorable conformation is seen in solution structure ($\phi \approx 60°, \psi \approx -75°$), except for one conformer (Blanco et al. 1997). The built-in strain apparently triggers local motions at this site. Both N47 and D48 display sharply elevated B-factors and several of the side-chain atoms in D48 cannot be resolved by x-ray crystallography at all. D48 is also the only residue featuring exchange broadening in solution ($R_{ex} = 28$ s^{-1} at 600 MHz) (Chevelkov et al. 2010). As it turns out, the RDC data from N47 and D48 can be accommodated using the existing structural models. Specifically, if these two residues are 'grafted' from the NMR structure 1AEY onto the crystallographic structure 1U06, some of the resulting hybrid models lead to a satisfactory RDC fit.

Alignment

The alignment parameters derived from multiple SH3 samples are summarized in Tables 1, 2 and in Fig. 7. As already discussed, special care was taken to ensure that the alignment conditions are the same throughout each series of samples. In the case of $r = 0.85$ samples, the residual quadrupolar splittings $\Delta(^2\text{H})$ proved to be fairly uniform and reproducible (with the apparent exception of dSH3-*ml*-dSH3, see Table 1). For $r = 0.96$, the splittings were lower than expected and varied significantly from one sample to the other (see Table 2). This latter media appears to be only marginally stable at 25°C (Rückert and Otting 2000). Generally, the dilute PEG/hexanol media used in NMR experiments falls close to the boundary between lamellar

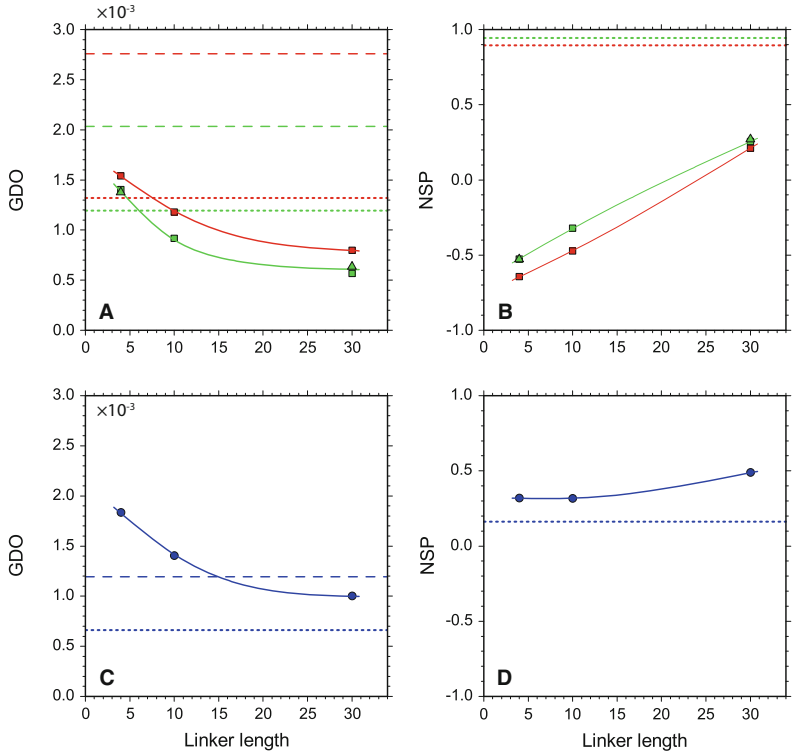


Fig. 7 GDO and NSP alignment parameters in a series of constructs of α-spc SH3 domain. **a** and **b** Experimental data from dSH3-*sl*-dSH3, dSH3-*ml*-dSH3, and dSH3-*ll*-dSH3 tandem constructs (*squares* for the main series of samples, *triangles* for replicate samples), as well as dSH3 and fSH3 single-domain constructs (*dashed* and *dotted horizontal lines*, respectively). The data from $r = 0.85$ and 0.96 samples are indicated in *green* and *red*, respectively. Spline curves are added to guide the eye. The NSP parameter is calculated relative to the dSH3 construct; the corresponding value for the dSH3 itself is 1.0. **c** and **d** The corresponding PALES-based simulations using the steric alignment model which assumes planar bicelle bilayers with the thickness 28.6 Å

phase L_α and a 'sponge phase' L_3, which itself has a complex and variable morphology (Gaemers and Bax 2001). In this situation, minor variations in the composition of the sample—possibly, the differences between the protein constructs themselves—may cause significant changes in the alignment. To account for these changes, we choose to normalize the GDO parameter according to the magnitude of $\Delta(^2H)$, see Table 1. In doing so we assume that $\Delta(^2H)$ reflects the aligning properties of the media not only with respect to water, but also to different α-spectrin SH3 constructs. This is an admittedly *ad hoc* approach, with only limited experimental justification (Rückert and Otting 2000). Note, however, that it appears justified for a series of samples where the alignment is dictated by the (invariant) properties of the α-spc SH3 domain. At least in the case of $r = 0.85$ media the alignment conditions seem to be highly reproducible from one sample to the other.

The consistency of different sample preparations can be verified through measurements on replicate samples. The results from the duplicate dSH3-*sl*-dSH3 and dSH3-*ll*-dSH3 samples are in very good agreement with each other (see Table 1; also compare squares and rectangles in Fig. 7). Furthermore, the data taken at $r = 0.85$ and 0.96 are also in good agreement insofar as rhombicity and the orientation of the alignment frame are concerned. One apparent exception is dSH3-*ml*-dSH3, cf. Tables 1 and 2. This construct displays the rhombicity R approaching the maximum value of 2/3. Under these circumstances, the labeling of the principal axes of the alignment tensor becomes ambiguous, causing apparent differences in the Euler angles. In fact, the two data sets are highly consistent, as confirmed by the NSP of the respective alignment tensors, 0.998.

The only sample where reproducibility truly appears to be an issue is dSH3. While the orientation of the alignment

frame is well reproduced, the rhombicity shows a significant amount of variation, Tables 1 and 2. Furthermore, the GDO value varies significantly from one preparation to the other. One thing is certain, however: the alignment of dSH3 is unexpectedly strong in all anisotropic samples, cf. Fig. 6 (also confirmed by the results from an additional dSH3 sample, not shown). In what follows, this observation is discussed in greater detail.

Figure 7a provides the experimental verification of the concept presented in this paper, namely, that the tandem with a short linker should align strongly, whereas the tandem with a long linker should align more weakly (given that the shape of the domains is close to spherical, the linker is unstructured, and there is no significant domain–domain interactions). Furthermore, it appears that the simple steric alignment model, as implemented in PALES, in conjunction with the conformational ensembles generated with the help of TraDES, provide an adequate explanation for the experimental observations (compare squares/triangles in Fig. 7a and circles in Fig. 7c). Note that the PALES-based prediction algorithm does not involve any adjustable parameters (see Materials & Methods). The NSP data in Fig. 7b confirm our hypothesis that with increase in linker length the tandem domains begin to behave as independent entities, i.e. align similarly to isolated domains.

From inspection of Fig. 7 it becomes obvious, however, that the steric alignment model is only partially successful in reproducing the experimental data. While the GDO values for tandem constructs are predicted with reasonable accuracy, the predictions for isolated domains, dSH3 and fSH3, are off the mark (cf. dashed/dotted lines in Fig. 7a, c). The case of the dSH3 is particularly instructive. This domain represents a spheroid with modest anisotropy, $I_{\parallel}/I_{\perp} = 0.83$. Its structure is well defined: only one N-terminal and one C-terminal residue are classified as 'unstructured'. Assuming that the mechanism of alignment is steric, dSH3 should show only a modest degree of alignment. Yet the experimental GDO value is exceedingly high (see also Fig. 6a). Moreover, the experimentally determined alignment frame bears no similarity to the predicted one, as indicated by the low NSP value, −0.11. This leads us to suggest that the alignment mechanism in the PEG/hexanol media is, in fact, different from the simple steric model postulated in PALES.

To further investigate this possibility, we turn to the literature reports of RDC measurements in the PEG/hexanol media. For ubiquitin, the orientation of the alignment frame is predicted by PALES with reasonable accuracy, NSP = 0.78–0.86. However, the GDO values are overestimated by a factor 2.1–4.5 (Lakomek et al. 2006; Briggman and Tolman 2003). Our own measurements on ubiquitin in PEG/hexanol media at pH 3.5 led to similar results, with GDO overestimated by 2.1–2.3 (data not shown). The agreement becomes even poorer if the results are scaled according to $\Delta(^2H)$. The data from another small protein, GB3 domain, produce NSP = 0.94 and another poor GDO prediction, off by a factor 3.6 (Ulmer et al. 2003). Taken together with our current data for α-spectrin SH3, these results suggest that PALES can both *under*- and *overestimate* the degree of alignment by a wide margin.

This shortcoming seems to be specific to the PEG/hexanol media. In the case of DMPC/DHPC bicelles, the steric PALES model fares much better. For ubiquitin, NSP = 0.96–0.97 and the predicted GDO value is only 1.3 times higher than the experimental value. For GB3 domain, NSP = 0.90 and the predicted GDO is essentially exact. Clearly, it would have been desirable to repeat our experiments in the DMPC/DHPC bicelles. Unfortunately, fSH3 and especially dSH3 constructs tend to precipitate at pH 6.5–7.0 required by this alignment media (van Rossum et al. 2003).

Alignment mechanisms

Unexpectedly, the results Fig. 7a indicate that nearly-spherical dSH3 domain aligns stronger than the markedly anisotropic dSH3-*sl*-dSH3 construct. In this section we discuss the potential sources of this behavior. While we cannot offer any single definitive explanation, it is useful to review various factors contributing to PEG/hexanol alignment.

The result in Fig. 7 appears paradoxical only in the light of the simple steric model involving planar bicelles. Alternatively, if the morphology of the media is complex, then a steric mechanism can, in principle, lead to an arbitrary alignment tensor. Indeed, assume for the sake of argument that anisotropic media contains cavities that can (transiently) accommodate certain proteins. Assume further that the cavities are shaped such that they form 'lock and key' pair with a given protein. Clearly, the resulting alignment tensor will be different from the one that is predicted by PALES.

In fact, recent studies showed that the topology of the most used liquid-crystal media is far more complex than previously believed. The DMPC/DHPC bicelles, which were originally described as large disks, have been instead visualized as "two-dimensional networks of flattened, highly branched, cylindrical micelles and lamellar sheets perforated by large irregular holes" (van Dam et al. 2006). Similar observations have been made with regard to the PEG-hexanol media. Early on it has been noted that the L_α bilayers display collective wave-like modulation (undulation) (Strey et al. 1990; Jonstromer and Strey 1992). Gaemers and Bax pointed out that dilute conditions used in

NMR experiments bring the system close to the transitions region where the lamellar phase L_α coexists with a 'sponge phase' L_3 (Gaemers and Bax 2001). Their diffusion data further suggest that the motion of probe molecules between the oriented bilayers (i.e. along the z-axis in the NMR tube) is significantly obstructed. These obstructions may arise from the 'bridges' between the adjacent bilayers, the presence of small globular structures (e.g. micelles) in the interbilayer space, etc. None of this potentially complex topology is taken into consideration in the PALES model—which may explain its apparent lack of accuracy, Fig. 7.

From a more general perspective, the failure of PALES can be attributed to site-specific interactions between the proteins and the media. Generally, PEG-based media are protein-friendly in a sense that there is little protein adsorption. Nevertheless, specific interactions involving PEG-hexanol have been observed for short peptides (Ohnishi and Shortle 2003), N-terminal domain of the enzyme I from bacterial phosphotransferase system (Suh et al. 2008), two cytoplasmic domains of chitobiose transporter (Jung et al. 2010), and folding intermediate of FF domain from HYPA/FBP11 (Korzhnev et al. 2010). There is no doubt that many more cases remain undocumented (dismissed as unsuccessful attempts to prepare aligned samples).

PEG doped with hexanol forms a bilayer where the surface is lined with hydroxyl headgroups. The interactions of the proteins with such bilayers have been modeled in several influential theoretical studies (Jeon et al. 1991; McPherson et al. 1998). All-atom force field simulations have been also reported (Lim and Herron 1992; Zheng et al. 2004; Zheng et al. 2005). These simulations clearly identify the preferred orientation of the protein (lysozyme) on a surface of a PEG bilayer. Of note, the binding interface appears to be formed by charged/polar residues interacting with PEG oxygens at or near the surface of the bilayer (Zheng et al. 2004). Of course, the binding is only weakly specific since many sites on the protein surface can form such favorable contacts. Note that electrostatic effects play a prominent role in molecular alignment even when the media is not charged—it is sufficient that the media is polar. These aspects has been extensively discussed in the small-molecule studies (Emsley et al. 1991; Terzis and Photinos 1994; Syvitski and Burnell 2000; Dingemans et al. 2003; Wardle et al. 2004).

To elucidate a possible role of electrostatics in PEG/hexanol media, we have prepared an additional sample of fSH3 containing 100 mM NaCl. If the alignment is purely steric, as is commonly held, then one would not expect to observe any significant changes upon addition of NaCl. In reality, however, the GDO parameter jumps from 1.19×10^{-3} to 1.85×10^{-3}, see Table 1.[1] This result is unexpected and, at a first glance, counterintuitive. Indeed, if electrostatic interactions are involved, then addition of salt should screen out these interactions and thus *lower* the degree of alignment. In fact, a number of plausible explanations can be suggested for this finding. The increase in GDO may result from the interplay between the electrostatic and steric mechanisms. Furthermore, it may also occur in the context of purely electrostatic alignment. As described by Wu and co-workers (Wu et al. 2006), excessive electrostatic repulsion drives the solute away from the charged media and into the bulk, where it cannot effectively align. Adding NaCl to the solution reverses this process, thus increasing the degree of alignment. Details of alignment notwithstanding, the observed dependence of the GDO on the ionic strength of the solution suggests that an electrostatic mechanism plays a role.

To further probe the relevance of electrostatic interactions, we have undertaken an additional series of PALES simulations. Unlike the steric model, which has no adjustable parameters, the electrostatic model contains several such quantities. As described in the Materials & Methods section, the ionic strength of the solution was set to the actual experimental value, 17 mM. The other two parameters, pH and the surface charge density of the medium σ, were optimized in an *ad hoc* fashion to obtain the best possible agreement with the five experimental GDO values (listed in Table 1). The optimization was conducted by means of the grid-search in the space of two parameters. Calculations at each point on the grid involved multiple PALES jobs (addressing five conformational ensembles, each comprised of several thousand conformers).

The optimized electrostatic model produced the predictions for GDO and NSP as shown in Fig. 8. The main result of this simulation is that it is indeed possible to recreate the situation where dSH3 and fSH3 display a higher degree of alignment than tandem constructs, as observed experimentally (see Fig. 7a). The parameters obtained from the model optimization are $\sigma = 0.05$ e/nm^2, effective pH 4.8. The value 0.05 is relatively low in comparison, for example, with Pf1 phage, -0.47 e/nm^2. In principle, surface charge may arise from adsorption of ions on the interface between water and non-ionic surfactants (Elworthy et al. 1971; Balzer 1993; Karraker and Radke 2002; Stubenrauch and von Klitzing 2003; Wardle et al. 2004). However, it appears that the charge density on the surface of a C12E5 bilayer in water is actually much lower than 0.05 e/nm^2 and negative (Balzer 1993; Stubenrauch and von Klitzing 2003), even in the presence of a small amount of ionic surfactant (Schomacker and Strey 1994; von Berlepsch and de Vries 2000). At the same time, the obtained effective pH value, 4.8 is significantly higher than the actual experimental pH, 3.5.

[1] In contrast, we found that alignment of ubiquitin in PEG/hexanol is insensitive to salt.

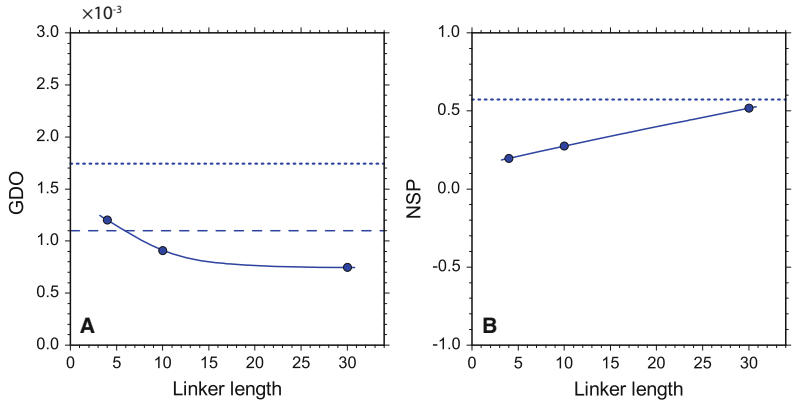


Fig. 8 The GDO and NSP alignment parameters for α-spc SH3 constructs as obtained from PALES-based electrostatic simulations. The optimized model assumes planar bicelle bilayers with surface charge density $\sigma = 0.05$ e/nm^2, and effective sample pH 4.8

In judging the results of these simulations one should keep in mind that our model of electrostatic alignment is rather crude. Specifically, the pK_a values used in the calculations have a typical error of 0.5–1 unit, with maximum error of up to 2 units (Li et al. 2005). Furthermore, the predictions of electrostatic alignment are highly sensitive to fine details of the charge distribution. For instance, the predicted alignment parameters may change in response to a conformational jump of a single charged side chain. To demonstrate this effect, we regenerated the SH3 conformational ensembles using the BAAAA structure of the SH3 domain, i.e. a structure where the single side chain, E17, is placed in an alternative conformation (see Materials & Methods). This minor structural modification led to significant decrease in the predicted GDO of the fSH3 sample, from 1.74×10^{-3} to 1.36×10^{-3}.

In conclusion, the presented PALES-based electrostatic simulation demonstrates the feasibility of the scenario in which dSH3 and fSH3 align more efficiently than the tandem constructs. Beyond that, we do not attach any particular significance to the results shown in Fig. 8. One can envisage that more sophisticated prediction tools will be developed in future, where a fully atomistic approach will be extended to the liquid crystal media. It can also be expected that MD-based methods will eventually emerge. The examples of such developments have been cited above (Dingemans et al. 2003; Wardle et al. 2004; Zangi and Engberts 2005; Zheng et al. 2004; Polyansky et al. 2005; Zheng et al. 2005; Aroulanda et al. 2006). The progress in this area, however, is limited by our relatively poor knowledge of the liquid crystal structure and dynamics.

Conclusion

In this study we focus on proteins comprised of two modular domains connected through a random-coil linker (α-spc SH3 domains connected through glycine-serine linker). When the linker is short, the tandem SH3 construct represents a dumbbell-shaped molecule with relatively little domain–domain mobility. Being placed in steric alignment media, such as PEG/hexanol, this molecule is expected to display a high degree of alignment. As the length of the linker is increased, the domains become effectively uncoupled and start behaving as independent entities. Consequently, the degree of alignment observed in each of the tandem domains approaches that of the isolated SH3 domain.

In qualitative terms, it has been noted before that differential domain alignment conveys information about interdomain mobility (Braddock et al. 2001; Ulmer et al. 2002; Poon et al. 2007; Holland et al. 2008; Wang et al. 2009). Here we attempted to quantify the dependence of the alignment parameters (amplitude and orientation) on the length of the linker. The form of the corresponding transition curves depends on the details of the system, e.g. the shape of the individual domains and the linker attachment points. In the case of electrostatic alignment, the trend can be easily reversed, i.e. the degree of alignment may increase with lengthening of the linker. The transition phenomenon itself, however, is sufficiently general.

To translate the proposed intuitive model into quantitative predictions, we generated a series of conformational ensembles representing α-spc SH3 constructs with different linker lengths (the disordered linkers and terminal sequences

were built using the program TraDES). These ensembles were used to predict the parameters of steric alignment (program PALES). To test the validity of the theoretical predictions, we prepared three tandem constructs of α-spc SH3 and two variants of the isolated SH3 domain. These chimeras were used to measure backbone $^{1}H^{N}$-^{15}N RDCs in two series of PEG/hexanol samples ($r = 0.85$ and 0.96). It was found that the alignment of tandem proteins is in line with expectations and agrees reasonably well with the results from PALES-based simulations. The isolated domains, however, displayed much stronger alignment than expected. Furthermore, the degree of alignment turned out to be sensitive to the ionic strength of the solution, which is generally not expected to be the case for a well-behaved steric alignment media.

Several factors may potentially contribute to this situation. First, the topology of the liquid crystal media is known to be more complex than envisaged in the simple PALES model (i.e. planar bilayers). Furthermore, there is a possibility that the phase equilibrium of the liquid crystal changes in response to addition of protein or salt (these changes may not necessarily be fully reflected in $\Delta(^{2}H)$) (Ottiger and Bax 1998). It is also likely that protein alignment in PEG/hexanol significantly depends on (weak) site-specific interactions. In particular, an electrostatic mechanism appears to play a role, visibly affecting the alignment of isolated domains.

In summary, it seems that PEG/hexanol bears little resemblance to an idealized alignment media envisioned in PALES. One would be well-advised to keep this in mind, especially in the context of studies where PALES is used to simulate RDC data from disordered proteins dissolved in PEG-based media (Alexandrescu and Kammerer 2003; Bernado et al. 2005; Jensen et al. 2008; Pinheiro et al. 2011). Indeed, if we cannot reliably predict the alignment of globular proteins, the predictions should be even less reliable for disordered proteins. In this latter case, all residues can potentially form site-specific contacts with the media and, furthermore, these contacts may significantly skew the original conformational equilibrium (for those protein molecules that are in the vicinity of the media and give rise to observable RDCs).

The chimera protein investigated in this work is comprised of two non-interacting domains connected by a variable-length random linker. It can thus be described as an example of the system with *minimal domain coupling*. This example provides an important point of reference for future studies of the proteins with substantial domain coupling, either through domain–domain interactions (e.g. Crk adaptor protein (Kobashigawa et al. 2007), Syk kinase (Zhang et al. 2008)) or through a structured linker (e.g. calmodulin (Ikura et al. 1991), troponin (Slupsky et al. 1995)). In the case of popular PEG/hexanol media, the usefulness of this approach is presently limited by lack of precise information about the morphology of the oriented phase and the details of alignment mechanism. We expect that the DMPC/DHPC media offers a better chance of success, although it is far more restrictive in terms of sample conditions. Further progress in this area depends on development of more accurate alignment prediction tools. Ultimately, such tools should be based on all-atom MD simulations modeling the interactions between the protein and the alignment media.

Acknowledgments This work has been funded through NSF grant MCB-044563. We thank Ryan Muir for his help with early versions of the structure-generating scripts. We are also grateful to Etti Harms, Nina Gorenstein, Josh Ward, and Yi Xue for their advice on different aspects of experimental work.

References

Akakura S, Kar B, Singh S, Cho L, Tibrewal N, Sanokawa-Akakura R, Reichman C, Ravichandran KS, Birge RB (2005) C-terminal SH3 domain of Crkll regulates the assembly and function of the DOCK180/ELMO Rac-GEF. J Cell Physiol 204:344–351

Alexandrescu AT, Kammerer RA (2003) Structure and disorder in the ribonuclease S-peptide probed by NMR residual dipolar couplings. Protein Sci 12:2132–2140

Andre I, Linse S, Mulder FAA (2007) Residue-specific pK_a determination of lysine and arginine side chains by indirect ^{15}N and ^{13}C NMR spectroscopy: Application to apo calmodulin. J Am Chem Soc 129:15805–15813

Aroulanda C, Celebre G, De Luca G, Longeri M (2006) Molecular ordering and structure of quasi-spherical solutes by liquid crystal NMR and Monte Carlo simulations: the case of norbornadiene. J Phys Chem B 110:10485–10496

Bae SH, Dyson HJ, Wright PE (2009) Prediction of the rotational tumbling time for proteins with disordered segments. J Am Chem Soc 131:6814–6821

Balzer D (1993) Cloud point phenomena in the phase behavior of alkyl polyglucosides in water. Langmuir 9:3375–3384

Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pK_a values for protein-ligand complexes. Proteins 73:765–783

Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12:1–16

Bax A, Tjandra N (1997) High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J Biomol NMR 10:289–292

Bax A, Kontaxis G, Tjandra N (2001) Dipolar couplings in macromolecular structure determination. Method Enzymol 339:127–174

Berlin K, O'Leary DP, Fushman D (2009) Improvement and analysis of computational methods for prediction of residual dipolar couplings. J Magn Reson 201:25–33

Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 102:17002–17007

Bertini I, Del Bianco C, Gelis I, Katsaros N, Luchinat C, Parigi G, Peana M, Provenzani A, Zoroddu MA (2004) Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. Proc Natl Acad Sci USA 101:6841–6846

Bertini I, Gupta YK, Luchinat C, Parigi G, Peana M, Sgheri L, Yuan J (2007) Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. J Am Chem Soc 129:12786–12794

Bewley CA (2001) Solution structure of a cyanovirin-N: Manα1-2Manα complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. Structure 9:931–940

Blanco FJ, Ortiz AR, Serrano L (1997) ^{1}H and ^{15}N NMR assignment and solution structure of the SH3 domain of spectrin: comparison of unrefined and refined structure sets with the crystal structure. J Biomol NMR 9:347–357

Braddock DT, Cai ML, Baber JL, Huang Y, Clore GM (2001) Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. J Am Chem Soc 123:8634–8635

Braddock DT, Baber JL, Levens D, Clore GM (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP KKH3 and single-stranded DNA. EMBO J 21:3476–3485

Briggman KB, Tolman JR (2003) De Novo determination of bond orientations and order parameters from residual dipolar couplings with high accuracy. J Am Chem Soc 125:10164–10165

Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30:1545–1614

Brosey CA, Chagot ME, Ehrhardt M, Pretto DI, Weiner BE, Chazin WJ (2009) NMR analysis of the architecture and functional remodeling of a modular multidomain protein, RPA. J Am Chem Soc 131:6346–6347

Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 84: 7524–7528

Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins Struct Funct Genet 21:167–195

Cesareni G, Gimona M, Sudol M, Yaffe M (eds) (2005) Modular protein domains. Wiley-VCH, Weinheim

Chen K, Tjandra N (2008) Extended model free approach to analyze correlation functions of multidomain proteins in the presence of motional coupling. J Am Chem Soc 130:12745–12751

Chevelkov V, Faelber K, Diehl A, Heinemann U, Oschkinat H, Reif B (2005) Detection of dynamic water molecules in a microcrystalline sample of the SH3 domain of α-spectrin by MAS solid-state NMR. J Biomol NMR 31:295–310

Chevelkov V, Zhuravleva AV, Xue Y, Reif B, Skrynnikov NR (2007) Combined analysis of ^{15}N relaxation data from solid- and solution-state NMR spectroscopy. J Am Chem Soc 129: 12594–12595

Chevelkov V, Xue Y, Linser R, Skrynnikov NR, Reif B (2010) Comparison of solid-state dipolar couplings and solution relaxation data provides insight into protein backbone dynamics. J Am Chem Soc 132:5015–5017

Cipres A, Abassi YA, Vuori K (2007) Abl functions as a negative regulator of Met-induced cell motility via phosphorylation of the adapter protein CrkII. Cell Signal 19:1662–1670

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120:6836–6837

de la Torre JG, Huertas ML, Carrasco B (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys J 78:719–730

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe—a multidimensional spectral processing system based on unix pipes. J Biomol NMR 6:277–293

Denisov AY, Kloser E, Gray DG, Mittermaier AK (2010) Protein alignment using cellulose nanocrystals: practical considerations and range of application. J Biomol NMR 47:195–204

Dingemans T, Photinos DJ, Samulski ET, Terzis AF, Wutz C (2003) Ordering of apolar and polar solutes in nematic solvents. J Chem Phys 118:7046–7061

Dominy BN, Brooks CL (1999) Development of a generalized Born model parametrization for proteins and nucleic acids. J Phys Chem B 103:3765–3773

Donaldson LW, Gish G, Pawson T, Kay LE, Forman-Kay JD (2002) Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. Proc Natl Acad Sci USA 99:14053–14058

Elworthy PH, Rogers JA, Florence AT (1971) Stabilization of oil-in-water emulsions by nonionic detergents. 5. Effect of salts on rates of coalescence in a chlorobenzene emulsion. J Colloid Interf Sci 35:23–33

Emsley JW, Palke WE, Shilstone GN (1991) The inclusion of electrostatic and dispersion interactions into potentials of mean torque for solutes dissolved in uniaxial liquid-crystal solvents. Liq Cryst 9:643–648

Feldman HJ, Hogue CWV (2000) A fast method to sample real protein conformational space. Proteins Struct Funct Genet 39:112–131

Feller SM (2001) Crk family adaptors—signalling complex formation and biological roles. Oncogene 20:6348–6371

Fischer D, Geyer A (2005) NMR spectroscopic characterization of the membrane affinity of polyols. Magn Reson Chem 43:893–901

Fitzkee NC, Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci USA 101:12497–12502

Freyssingeas É, Nallet F, Roux D (1996) Measurement of the membrane flexibility in lamellar and ''sponge'' phases of the $C_{12}E_5$/hexanol/water system. Langmuir 12:6028–6035

Futterer K, Wong J, Grucza RA, Chan AC, Waksman G (1998) Structural basis for syk tyrosine kinase ubiquity in signal transduction pathways revealed by the crystal structure of its regulatory SH2 domains bound to a dually phosphorylated ITAM peptide. J Mol Biol 281:523–537

Gaemers S, Bax A (2001) Morphology of three lyotropic liquid crystalline biological NMR media studied by translational diffusion anisotropy. J Am Chem Soc 123:12343–12352

Gelis I, Bonvin AMJJ, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell 131:756–769

Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A (2005) H++: a server for estimating pK_as and adding missing hydrogens to macromolecules. Nucl Acids Res 33:W368–W371

Hakansson KO (2009) The structure of Mg-ATPase nucleotide-binding domain at 1.6 angstrom resolution reveals a unique ATP-binding motif. Acta Crystallogr D 65:1181–1186

Heikkinen O, Permi P, Koskela H, Ylanne J, Kilpelainen I (2009) ^{1}H, ^{13}C and ^{15}N resonance assignments of the human filamin A tandem immunoglobulin-like domains 16–17 and 18–19. Biomol NMR Assign 3:53–56

Holland NB, Nishimiya Y, Tsuda S, Sonnichsen FD (2008) Two domains of RD3 antifreeze protein diffuse independently. Biochemistry 47:5935–5941

Huston JS, Levinson D, Mudgetthunter M, Tai MS, Novotny J, Margolies MN, Ridge RJ, Bruccoleri RE, Haber E, Crea R, Oppermann H (1988) Protein engineering of antibody-binding sites: recovery of specific activity in an anti-digoxin single-chain

Fv analog produced in *Escherichia Coli*. Proc Natl Acad Sci USA 85:5879–5883

Ikegami T, Verdier L, Sakhaii P, Grimme S, Pescatore B, Saxena K, Fiebig KM, Griesinger C (2004) Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions. J Biomol NMR 29:339–349

Ikura M, Kay LE, Krinks M, Bax A (1991) Triple-resonance multidimensional NMR study of calmodulin complexed with the binding domain of skeletal muscle Myosin Light-Chain Kinase: indication of a conformational change in the central helix. Biochemistry 30:5498–5504

Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RWH, Blackledge M (2008) Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: Application to the molecular recognition element of Sendai virus nucleoprotein. J Am Chem Soc 130:8055–8061

Jeon SI, Lee JH, Andrade JD, De Gennes PG (1991) Protein surface interactions in the presence of polyethylene oxide. I. Simplified theory. J Colloid Interf Sci 142:149–158

Jonstromer M, Strey R (1992) Nonionic bilayers in dilute solutions: effect of additives. J Phys Chem 96:5993–6000

Jung YS, Cai ML, Clore GM (2010) Solution structure of the $IIA^{Chitobiose}$-$IIB^{Chitobiose}$ complex of the N,N'-diacetylchitobiose branch of the *Escherichia Coli* phosphotransferase system. J Biol Chem 285:4173–4184

Karraker KA, Radke CJ (2002) Disjoining pressures zeta potentials and surface tensions of aqueous non-ionic surfactant/electrolyte solutions: theory and comparison to experiment. Adv Colloid Interfac 96:231–264

Keizers PHJ, Desreux JF, Overhand M, Ubbink M (2007) Increased paramagnetic effect of a lanthanide protein probe by two-point attachment. J Am Chem Soc 129:9292–9293

Khandogin J, Brooks CL (2006) Toward the accurate first-principles prediction of ionization equilibria in proteins. Biochemistry 45:9363–9373

Kobashigawa Y, Sakai M, Naito M, Yokochi M, Kumeta H, Makino Y, Ogura K, Tanaka S, Inagaki F (2007) Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. Nat Struct Mol Biol 14:503–510

Koenig BW, Hu JS, Ottiger M, Bose S, Hendler RW, Bax A (1999) NMR measurement of dipolar couplings in proteins aligned by transient binding to purple membrane fragments. J Am Chem Soc 121:1385–1386

Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE (2010) A transient and low-populated protein-folding intermediate at atomic resolution. Science 329:1312–1316

Lakomek NA, Carlomagno T, Becker S, Griesinger C, Meiler J (2006) A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. J Biomol NMR 34:101–115

Lazar GA, Desjarlais JR, Handel TM (1997) De novo design of the hydrophobic core of ubiquitin. Protein Sci 6:1167–1178

Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pK_a values. Proteins 61:704–721

Lim K, Herron JN (1992) Molecular simulation of protein-PEG interactions. In Harris JM (ed) Poly(ethylene glycol) chemistry: biotechnical and biomedical applications. Plenum Press, New York, pp 29–56

Linghu H, Tsuda M, Makino Y, Sakai M, Watanabe T, Ichihara S, Sawa H, Nagashima K, Mochizuki N, Tanaka S (2006) Involvement of adaptor protein Crk in malignant feature of human ovarian cancer cell line MCAS. Oncogene 25:3547–3556

Maltsev AS, Ahmed AH, Fenwick MK, Jane DE, Oswald RE (2008) Mechanism of partial agonism at the GluR2 AMPA receptor: Measurements of lobe orientation in solution. Biochemistry 47:10600–10610

Martinez JC, Pisabarro MT, Serrano L (1998) Obligatory steps in protein folding and the conformational diversity of the transition state. Nat Struct Biol 5:721–729

McPherson T, Kidane A, Szleifer I, Park K (1998) Prevention of protein adsorption by tethered poly(ethylene oxide) layers: experiments and single-chain mean-field analysis. Langmuir 14:176–186

Miller CT, Chen G, Gharib TG, Wang H, Thomas DG, Misek DE, Giordano TJ, Yee J, Orringer MB, Hanash SM, Beer DG (2003) Increased c-Crk proto-oncogene expression is associated with an aggressive phenotype in lung adenocarcinomas. Oncogene 22:7950–7957

Mohana-Borges R, Goto NK, Kroon GJA, Dyson HJ, Wright PE (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. J Mol Biol 340: 1131–1142

Moltke S, Grzesiek S (1999) Structural constraints from residual tensorial couplings in high resolution NMR without an explicit term for the alignment tensor. J Biomol NMR 15:77–82

Moreno-Murciano MP, Monleon D, Marcinkiewcz C, Calvete JJ, Celda B (2003) NMR solution structure of the non-RGD disintegrin obtustatin. J Mol Biol 329:135–145

Muralidharan V, Dutta K, Cho J, Vila-Perello M, Raleigh DP, Cowburn D, Muir TW (2006) Solution structure and folding characteristics of the C-terminal SH3 domain of c-Crk-II. Biochemistry 45:8874–8884

Musacchio A, Noble M, Pauptit R, Wierenga R, Saraste M (1992) Crystal structure of a Src-homology 3 (SH3) domain. Nature 359:851–855

Nishihara H, Tanaka S, Tsuda M, Oikawa S, Maeda M, Shimizu M, Shinomiya H, Tanigami A, Sawa H, Nagashima K (2002) Molecular and immunohistochemical analysis of signaling adaptor protein Crk in human cancers. Cancer Lett 180:55–61

Ogawa S, Toyoshima H, Kozutsumi H, Hagiwara K, Sakai R, Tanaka T, Hirano N, Mano H, Yazaki Y, Hirai H (1994) The C-Terminal SH3 domain of the mouse c-Crk protein negatively regulates tyrosine phosphorylation of Crk associated p130 in rat 3Y1 cells. Oncogene 9:1669–1678

Ohnishi S, Shortle D (2003) Observation of residual dipolar couplings in short peptides. Proteins Struct Funct Genet 50:546–551

Opella SJ, De Angelis AA (2007) Bicelle samples for solid-state NMR of membrane proteins. Nat Protoc 2:2332–2338

Ottiger M, Bax A (1998) Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. J Biomol NMR 12:361–372

Ottiger M, Delaglio F, Bax A (1998) Measurement of *J* and dipolar couplings from simplified two-dimensional NMR spectra. J Magn Reson 131:373–378

Peterson ME, Long EO (2008) Inhibitory receptor signaling via tyrosine phosphorylation of the adaptor Crk. Immunity 29: 578–588

Pinheiro AS, Marsh JA, Forman-Kay JD, Peti W (2011) Structural Signature of the MYPT1-PP1 Interaction. J Am Chem Soc 133:73–80

Polyansky AA, Volynsky PE, Nolde DE, Arseniev AS, Efremov RG (2005) Role of lipid charge in organization of water/lipid bilayer interface: Insights via computer simulations. J Phys Chem B 109:15052–15059

Poon DKY, Withers SG, McIntosh LP (2007) Direct demonstration of the flexibility of the glycosylated proline-threonine linker in the Cellulomonas fimi xylanase Cex through NMR spectroscopic analysis. J Biol Chem 282:2091–2100

Prestegard JH, Al-Hashimi HM, Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. Q Rev Biophys 33:371–424

Rodriguez-Castaneda F, Haberz P, Leonov A, Griesinger C (2006) Paramagnetic tagging of diamagnetic proteins for solution NMR. Magn Reson Chem 44:S10–S16

Rückert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. J Am Chem Soc 122:7793–7797

Sarkar P, Reichman C, Saleh T, Birge RB, Kalodimos CG (2007) Proline cis-trans isomerization controls autoinhibition of a signaling protein. Mol Cell 25:413–426

Sass J, Cordier F, Hoffmann A, Rogowski M, Cousin A, Omichinski JG, Lowen H, Grzesiek S (1999) Purple membrane induced alignment of biological macromolecules in the magnetic field. J Am Chem Soc 121:2047–2055

Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S (2000) Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. J Biomol NMR 18:303–309

Schomacker R, Strey R (1994) Effect of ionic surfactants on nonionic bilayers: bending elasticity of weakly charged membranes. J Phys Chem 98:3908–3912

Schwieters CD, Clore GM (2001) Internal coordinates for molecular dynamics and minimization in structure determination and refinement. J Magn Reson 152:288–302

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73

Shakhnovich EI, Gutin AM (1989) Formation of unique structure in polypeptide chains. Theoretical investigation with the aid of a replica approach. Biophys Chem 34:187–199

Simon K, Xu J, Kim C, Skrynnikov NR (2005) Estimating the accuracy of protein structures using residual dipolar couplings. J Biomol NMR 33:83–93

Skrynnikov NR, Goto NK, Yang DW, Choy WY, Tolman JR, Mueller GA, Kay LE (2000) Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. J Mol Biol 295:1265–1273

Slupsky CM, Kay CM, Reinach FC, Smillie LB, Sykes BD (1995) Calcium-induced dimerization of troponin C: mode of interaction and use of trifluoroethanol as a denaturant of quaternary structure. Biochemistry 34:7365–7375

Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445:618–622

Strey R, Schomacker R, Roux D, Nallet F, Olsson U (1990) Dilute lamellar and L_3 phases in the binary water-C12E5 system. J Chem Soc Faraday T 86:2253–2261

Stubenrauch C, von Klitzing R (2003) Disjoining pressure in thin liquid foam and emulsion films—new concepts and perspectives. J Phys Condens Matt 15:R1197–R1232

Suh JY, Cai ML, Clore GM (2008) Impact of phosphorylation on structure and thermodynamics of the interaction between the N-terminal domain of enzyme I and the histidine phosphocarrier protein of the bacterial phosphotransferase system. J Biol Chem 283:18980–18989

Syvitski RT, Burnell EE (2000) Dipole-induced ordering in nematic liquid crystals. II. The elusive holy grail. J Chem Phys 113:3452–3465

Takino T, Nakada M, Miyamori H, Yamashita J, Yamada KM, Sato H (2003) CrkI adapter protein modulates cell migration and invasion in glioblastoma. Cancer Res 63:2335–2337

Terzis AF, Photinos DJ (1994) Electrostatic interactions in liquid crystals: ordering of rigid solutes in nematic solvents. Mol Phys 83:847–865

Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114

Tollinger M, Forman-Kay JD, Kay LE (2002) Measurement of side-chain carboxyl pK_a values of glutamate and aspartate residues in an unfolded protein by multinuclear NMR spectroscopy. J Am Chem Soc 124:5714–5717

Tolman JR, Al-Hashimi HM, Kay LE, Prestegard JH (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. J Am Chem Soc 123:1416–1424

Tzakos AG, Grace CRR, Lukavsky PJ, Riek R (2006) NMR techniques for very large proteins and RNAs in solution. Annu Rev Biophys Biomol Struct 35:319–342

Ulmer TS, Werner JM, Campbell ID (2002) SH3-SH2 domain orientation in Src kinases: NMR studies of Fyn. Structure 10:901–911

Ulmer TS, Ramirez BE, Delaglio F, Bax A (2003) Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. J Am Chem Soc 125:9179–9191

Valafar H, Prestegard JH (2004) REDCAT: a residual dipolar coupling analysis tool. J Magn Reson 167:228–241

van Dam L, Karlsson G, Edwards K (2006) Morphology of magnetically aligning DMPC/DHPC aggregates-perforated sheets, not disks. Langmuir 22:3280–3285

van Rossum BJ, Castellani F, Pauli J, Rehbein K, Hollander J, de Groot HJM, Oschkinat H (2003) Assignment of amide proton signals by combined evaluation of HN, NN and HNCA MAS-NMR correlation spectra. J Biomol NMR 25:217–223

Vega MC, Martínez JC, Serrano L (2000) Thermodynamic and structural characterization of Asn and Ala residues in the disallowed II' region of the Ramachandran plot. Protein Sci 9:2322–2328

von Berlepsch H, de Vries R (2000) Weakly charged lamellar bilayer system: interplay between thermal undulations and electrostatic repulsion. Eur Phys J E1:141–152

Walsh JD, Meier K, Ishima R, Gronenborn AM (2010) NMR studies on domain diffusion and alignment in modular GB1 repeats. Biophys J 99:2636–2646

Wang W, Weng J, Zhang X, Liu M, Zhang M (2009) Creating conformational entropy by increasing interdomain mobility in ligand binding regulation: a revisit to N-terminal tandem PDZ domains of PSD-95. J Am Chem Soc 131:787–796

Wardle KE, Carlson E, Henderson D, Rowley RL (2004) Molecular-dynamics simulation of the effect of ions on a liquid–liquid interface for a partially miscible mixture. J Chem Phys 120:7681–7688

Watanabe T, Tsuda M, Makino Y, Ichihara S, Sawa H, Minami A, Mochizuki N, Nagashima K, Tanaka S (2006) Adaptor molecule Crk is required for sustained phosphorylation of Grb2-associated binder 1 and hepatocyte growth factor-induced cell motility of human synovial sarcoma cell lines. Mol Cancer Res 4:499–510

Wong V, Case DA, Szabo A (2009) Influence of the coupling of interdomain and overall motions on NMR relaxation. Proc Natl Acad Sci USA 106:11016–11021

Wu B, Petersen M, Girard F, Tessari M, Wijmenga SS (2006) Prediction of molecular alignment of nucleic acids in aligned media. J Biomol NMR 35:103–115

Yang DW, Nagayama K (1996) A sensitivity-enhanced method for measuring heteronuclear long-range coupling constants from the displacement of signals in two 1D subspectra. J Magn Reson Ser A 118:117–121

Zangi R, Engberts JBFN (2005) Physisorption of hydroxide ions from aqueous solution to a hydrophobic surface. J Am Chem Soc 127:2272–2276

Zhang YB, Zuiderweg ERP (2004) The 70-kDa heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains. Proc Natl Acad Sci USA 101:10272–10277

Zhang YJ, Oh H, Burton RA, Burgner JW, Geahlen RL, Post CB (2008) Tyr130 phosphorylation triggers Syk release from antigen receptor by long-distance conformational uncoupling. Proc Natl Acad Sci USA 105:11760–11765

Zheng J, Li LY, Chen SF, Jiang SY (2004) Molecular simulation study of water interactions with oligo (ethylene glycol)-terminated alkanethiol self-assembled monolayers. Langmuir 20:8931–8938

Zheng J, Li LY, Tsao HK, Sheng YJ, Chen SF, Jiang SY (2005) Strong repulsive forces between protein and oligo (ethylene glycol) self-assembled monolayers: A molecular simulation study. Biophys J 89:158–166

Zweckstetter M (2006) Prediction of charge-induced molecular alignment: residual dipolar couplings at pH 3 and alignment in surfactant liquid crystalline phases. Eur Biophys J Biophys Lett 35:170–180

Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 3:679–690

Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. J Am Chem Soc 122: 3791–3792

Zweckstetter M, Hummer G, Bax A (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J 86:3444–3460

Journal of Magnetic Resonance 241 (2014) 155–169

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Proton-decoupled CPMG: A better experiment for measuring ^{15}N R_2 relaxation in disordered proteins

Tairan Yuwen, Nikolai R. Skrynnikov *

Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 27 June 2013
Revised 15 August 2013
Available online 23 August 2013

Keywords:
^{15}N relaxation
Intrinsically disordered proteins
CPMG experiment
Spin-lock experiment
Ubiquitin
α-Spectrin SH3 domain
Solvent exchange
Backbone dynamics
Arginine side-chain dynamics
Spin simulations

ABSTRACT

^{15}N R_2 relaxation is one of the most informative experiments for characterization of intrinsically disordered proteins (IDPs). Small changes in nitrogen R_2 rates are often used to determine how IDPs respond to various biologically relevant perturbations such as point mutations, posttranslational modifications and weak ligand interactions. However collecting high-quality ^{15}N relaxation data can be difficult. Of necessity, the samples of IDPs are often prepared with low protein concentration and the measurement time can be limited because of rapid sample degradation. Furthermore, due to hardware limitations standard experiments such as ^{15}N spin-lock and CPMG can sample the relaxation decay only to ca. 150 ms. This is much shorter than ^{15}N T_2 times in disordered proteins at or near physiological temperature. As a result, the sampling of relaxation decay profiles in these experiments is suboptimal, which further lowers the precision of the measurements.

Here we report a new implementation of the proton-decoupled (PD) CPMG experiment which allows one to sample ^{15}N R_2 relaxation decay up to ca. 0.5–1 s. The new experiment has been validated through comparison with the well-established spin-lock measurement. Using dilute samples of denatured ubiquitin, we have demonstrated that PD-CPMG produces up to 3-fold improvement in the precision of the data. It is expected that for intrinsically disordered proteins the gains may be even more substantial. We have also shown that this sequence has a number of favorable properties: (i) the spectra are recorded with narrow linewidth in nitrogen dimension; (ii) ^{15}N offset correction is small and easy to calculate; (iii) the experiment is immune to various spurious effects arising from solvent exchange; (iv) the results are stable with respect to pulse miscalibration and *rf* field inhomogeneity; (v) with minimal change, the pulse sequence can also be used to measure R_2 relaxation of $^{15}N^{\varepsilon}$ spins in arginine side chains. We anticipate that the new experiment will be a valuable addition to the NMR toolbox for studies of IDPs.

1. Introduction

^{15}N relaxation measurement is one of the most useful NMR experiments – it is fast, reasonably simple to interpret, and immediately provides information on protein backbone flexibility. This information can be further used to characterize the role of motions (i.e. conformational entropy) in ligand binding [1,2], protein stability [3,4], allosteric regulation [5,6], etc.

The spin context of ^{15}N relaxation is fairly straightforward. It is normally assumed that nitrogen relaxation is driven by $^{1}H^{N}$–^{15}N dipolar interaction and, to a lesser degree, by ^{15}N CSA mechanism. Nevertheless, even in this sufficiently simple spin system there is room for uncertainty – the appropriate length of N–H bond was a subject of debate and the degree of anisotropy and site-to-site variation of the ^{15}N CSA tensors have also been widely discussed [7,8]. Furthermore, small and difficult-to-quantify R_{ex} contributions into ^{15}N R_2 appear to be more common than previously thought [9].

In addition to these fundamental sources of uncertainty, there are also a number of errors associated with specific measurement schemes that have been uncovered in recent years. For example, it has been found that frequent application of ^{1}H pulses during the spin-lock period with the intention to suppress dipolar-CSA cross-correlation may actually re-introduce the undesirable J-coupling evolution [10]. In $^{1}H^{N}$–^{15}N saturation-transfer NOE experiment it was found that the standard saturation sequence tends to create a number of unwanted spin modes [11]. The artefacts associated with solvent exchange were also recently exposed [12].

Finally, there are also uncertainties associated with the choice of the model. The most common approach is to choose from several versions of the model-free model which make different assumptions about the underlying dynamics – in particular, with regard to local motions on ~1 ns time scale [13–15]. While there are statistical criteria that guide this choice, it has been shown that the outcome is often influenced by the random noise-like errors

* Corresponding author.
E-mail address: nikolai@purdue.edu (N.R. Skrynnikov).

1090-7807/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmr.2013.08.008

and depends on availability of data measured at different magnetic fields [16–18]. An additional level of complexity is added by the protein's overall tumbling, which should be properly modeled using a fully asymmetric diffusion tensor [19–22].

With all of these sources of uncertainty combined, it is not surprising that ^{15}N-relaxation-derived motional parameters are not necessarily very accurate. Consider, for example, the well-studied model protein ubiquitin, where several groups independently measured and analyzed ^{15}N relaxation in solution [20,23–25]. The backbone order parameters reported in these studies differ significantly from each other: the *rms* deviations between the independently determined sets of S^2 are in the range from 0.02 to 0.07 and the correlation coefficients are in the range from 0.56 to 0.94.

In practice, however, the perceived lack of accuracy is not very important so long as most experimental studies focus on the difference in order parameters, ΔS^2, rather than the absolute values. It is reasonably safe to assume that systematic biases subtract out in calculating ΔS^2 and the result correctly represents changes in local dynamics. The examples include changes in order parameters in response to ligand binding, point mutations, various post-translational modifications, etc. [26–29]. Given that such changes are typically small, on the order of 0.05 or less, the primary concern is the *precision* of the data – i.e. it is important to ensure that small ΔS^2 values can be experimentally reproduced and quantitated in a reliable fashion.

These considerations are especially relevant for intrinsically disordered proteins (IDPs), which have emerged as the area of major biomedical interest [30]. IDPs do not lend themselves to the standard methods of structural characterization. Instead, the dynamics-oriented experiments play a prominent role. In particular, ^{15}N relaxation and paramagnetic relaxation enhancements (PREs) are arguably two most informative experiments in the field of intrinsically disordered proteins. At the same time, it is often considerably more difficult to achieve high precision in measuring ^{15}N relaxation rates in IDPs. Many disordered proteins are prone to aggregation which makes it necessary to work with low-concentration NMR samples, often in the range 10–100 μM, and/or collect the data over a short period of time to avoid sample degradation [31–33]. If the studies are conducted at or near physiological conditions (37 °C, pH 7.4), the quality of the already crowded spectra suffers from the line-broadening caused by solvent exchange, which further lowers the signal-to-noise ratio. Finally, there is one other issue that is concerned specifically with ^{15}N R_2 measurements as discussed below.

For disordered proteins at 37 °C the ^{15}N T_2 relaxation times are relatively long, on the order of 0.5–1.0 s. On the other hand, the length of the spin-lock period τ_{rel} in the standard $R_{1\rho}$ experiment does not exceed ca. 100–150 ms, as dictated by hardware limitations. It is clearly not possible to properly sample the magnetization decay curve over such a short interval of time. This situation is illustrated in Fig. 1a, which shows the $R_{1\rho}$ decay curve from residue L15 in a 100 μM sample of denatured ubiquitin at 37 °C. The expansion of this relaxation profile in Fig. 1b demonstrates the problem with such measurement – generally, when τ_{rel} is much shorter than T_2 the decay profile appears flat, which makes it difficult to accurately determine the relaxation rate constant. A rigorous treatment indicates that the best results are obtained when τ_{rel} is extended beyond T_2 [34].

The restrictions concerning the maximum length of τ_{rel} are ultimately due to low gyromagnetic ratio of nitrogen, $\omega_1 = -\gamma_N B_1$. In order to generate sufficiently strong ^{15}N *rf* field which is necessary to suppress the effect of $^1J_{NH}$ coupling, a very strong current should be applied to the transmitter coil. This pushes the probe to the limits of its performance. In the case of spin-lock measurements, the room-temperature probe can sustain 1.7 kHz nitrogen *rf* field for a period of time ca. 150 ms. In the case of R_2 CPMG measurements, the probe can withstand the pulse train with the repetition time of 1 ms for ca. 100 ms. For early-generation cryogenic probes the power limits are even more restrictive. Extending τ_{rel} beyond these limits can lead to excessive heating of the sample and cause probe arcing [35] which severely degrades the performance of the pulse sequence. In the worst-case scenario, the probe can be permanently damaged.

It appears that the latest-generation hardware and particularly new cryogenic probes should make it possible to run spin-lock experiment with long τ_{rel} [36]. Preliminary data obtained by Ban and Lee suggest that nitrogen *rf* field with the strength of 3 kHz can be maintained for 500 ms using reasonable duty cycle of 20% (personal communication). In these trial experiments *rf* power remains stable (i.e. there is no observable amplifier drooping) and *rf* coil experiences little or no heating. The sample heating is also modest, even for high-salt samples, which can be attributed to the improvements in probe head design. These favorable properties can potentially lead to a robust R_2 experiment well suited for application to samples of IDPs at or near physiological conditions. However, to the best of our knowledge, such strategy has never been tested experimentally. One should also bear in mind that the latest-generation hardware which is necessary for such measurements is available only to a handful of research groups.

The problem can also be managed without recourse to advanced hardware. For example, Palmer and co-workers demonstrated that $^1J_{NH}$ coupling can be suppressed by using relatively weak ^{15}N spin lock plus a small number of strategically placed ^{1}H pulses [37]. This experiment, however, is semi-selective – to collect a complete $R_{1\rho}$ dataset one needs to repeat the measurements several times using different ^{15}N carrier settings.

The alternative solution, which we pursue in this report, is to use ^{1}H decoupling to suppress $^1J_{NH}$. Because of the high gyromagnetic ratio of the proton, it is easy to generate a sufficiently strong *rf* field without overloading the probe. In the presence of proton decoupling, ^{15}N 180° pulses can be applied only infrequently (for the purpose of refocusing chemical shift evolution). Using this approach, τ_{rel} can be extended by at least a factor of 5 without risking adverse consequences to a probe.

A suitable pulse sequence, proton-decoupled CPMG (PD-CPMG), has been introduced a long time ago and used to measure ^{15}N relaxation in proteins [38–41]. Essentially all of these reports note the presence of bias in R_2 rates measured with PD-CPMG, on the order of 5% [38,39,41]. Because of this unresolved issue PD-CPMG has seen little use over the last two decades.

Here we present a new and upgraded version of PD-CPMG which can reliably reproduce the results of the well-established $R_{1\rho}$ spin-lock experiment. The new pulse sequence allows one to extend the observation window τ_{rel} to 500 ms and beyond, Fig. 1c. The proper sampling of the relaxation decay curve, down to the level of 0.2 on the relative intensity scale, facilitates the accurate determination of R_2. For the samples at hand, the precision of the relaxation measurements is improved by up to a factor of 3 compared to the standard $R_{1\rho}$ experiment. In addition to the extended observation window, the new version of PD-CPMG experiment offers a number of other valuable properties, as discussed below.

2. Results

2.1. PD-CPMG pulse sequence for ^{15}N R_2 measurements

The new pulse sequence shown in Fig. 2 is essentially a decoupled, sensitivity-enhanced HSQC, similar to what has been previously used for ^{15}N $R_{1\rho}$ measurements [10]. Inserted in this sequence is the proton-decoupled CPMG element of the net

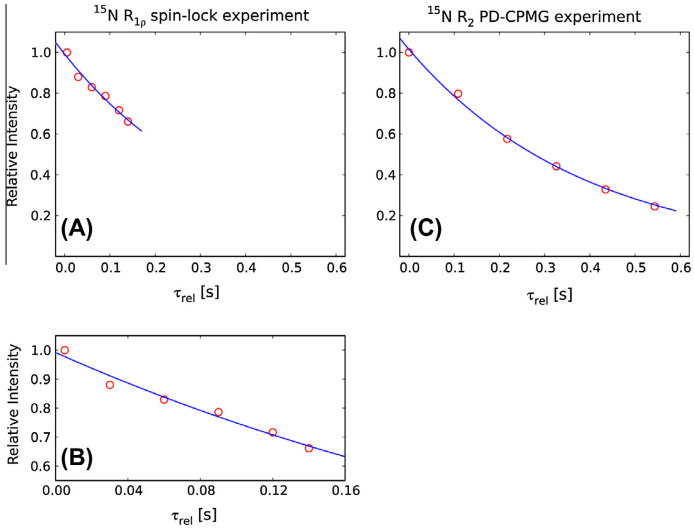


Fig. 1. Examples of relaxation curves from ^{15}N $R_{1\rho}$ spin-lock experiment (panels A and B) and R_2 PD-CPMG experiment (panel C). The data are from residue L15 in 100 μM sample of unfolded ubiquitin at 37 °C. Panel (B) is the expansion of panel (A).

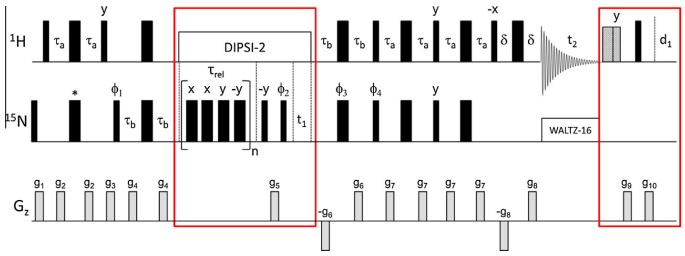


Fig. 2. PD-CPMG pulse sequence for measuring ^{15}N R_2 relaxation rates. The design is based on the experiment by Korzhnev et al. [10]; the portions of the sequence which are new and distinct are enclosed in red boxes. The pulse and delay settings listed below are for the measurements on denatured ubiquitin (see Section 5 for sample conditions). Hard ^{1}H pulses are applied with the carrier set on water resonance and *rf* field strength of 40 kHz. Two consecutive water purge pulses [51] have the durations of 6.0 and 3.7 ms and use the *rf* field of 12.4 kHz (shaded rectangles in the plot). DIPSI-2 decoupling [46] is applied with the carrier shifted to the middle of the amide region (8.3 ppm). The decoupling *rf* field corresponds to 90° pulse of duration 61 μs (*rf* field strength 4.1 kHz). Considering that one DIPSI-2 cycle corresponds to 10,360° rotation, the duration of the cycle is determined to be Δ_{DIPSI2} = 7.0218 ms. The ^{15}N CPMG sequence is applied in a form of $[\tau_{CP} - 180^\circ_x - \tau_{CP} - \tau_{CP} - 180^\circ_x - \tau_{CP} - \tau_{CP} - 180^\circ_y - \tau_{CP} - \tau_{CP} - 180^\circ_{-y} - \tau_{CP}]_n$ pulse train, where n is an integer number. When n = 0 this element becomes nil, i.e. it does not contain any pulses or delays. The interval between the consecutive 180° ^{15}N CPMG pulses should be adjusted according to the synchronization condition $2\tau_{CP} + p_{N(180)} = m\Delta_{DIPSI2}$, where $p_{N(180)}$ is the duration of 180° ^{15}N pulse and m is an integer number. In our main series of measurements we used the setting m = 2 corresponding to $2\tau_{CP}$ = 13.9676 ms. All ^{15}N pulses, including 180° CPMG pulses, have been applied with the maximum *rf* field strength, in our case 6.6 kHz ($p_{N(180)}$ = 76 μs). WALTZ-16 decoupling during the acquisition period employs 1.1 kHz *rf* field. The delay settings are τ_a = 2.25 ms, $\tau_b = 1/(4J_{NH})$ = 2.75 ms, and δ = 0.75 ms. The recycling delay employed is d_1 = 2 s. Gradient strengths in G/cm (length in ms) are g_1 = 5.0 (1.0), g_2 = 4.0 (0.5), g_3 = 10.0 (1.0), g_4 = 8.0 (0.5), g_5 = 25.0 (1.0), g_6 = 15.0 (1.25), g_7 = 4.0 (0.5), g_8 = 14.8 (0.125), g_9 = 20.0 (3.5), g_{10} = 20.0 (2.0). The phase cycle employed is ϕ_1 = (x, −x), ϕ_2 = y, ϕ_3 = 2(x), 2(y), 2(−x), 2(−y), ϕ_4 = x, ϕ_{rec} = (x, −x, −x, x), with DIPSI-2 *rf* field applied along x. Quadrature detection in F_1 is achieved by using the enhanced sensitivity approach [52,53] whereby separate spectra are recorded with (ϕ_4, g_6) and (ϕ_4 + π, −g_6) settings for each t_1 increment. Phase ϕ_2 is inverted in concert with the receiver phase for each t_1 increment such as to shift axial peaks to the edge of the spectrum [54]. While the sequence is primarily aimed at the backbone amide sites, it can also be adapted to measure arginine side-chain $^{15}N^\varepsilon$ R_2 relaxation rates. For this purpose ^{15}N carrier needs to be set to 85 ppm, rectangular ^{15}N 180° pulse marked by (*) should be replaced with r-SNOB pulse [55] with ca. 15 ppm bandwidth, and ^{15}N 180° pulses in the CPMG train need to be changed into REBURP shapes [56] with ca. 15 ppm bandwidth covering $^{15}N^\varepsilon$ resonances, but not $^{15}N^\eta$.

duration τ_{rel}. The CPMG pulse train uses [0013] phase cycle which was recently shown to have favorable compensation properties [42–45]. The pulse repetition rate in the PD-CPMG sequence is relatively low, with the pulse spacing $2\tau_{CP}$ ca. 14 ms. The proton decoupling during the CPMG train is achieved by application of the DIPSI-2 sequence [46] with the *rf* field strength 4 kHz. The relatively low level of combined ^{1}H and ^{15}N *rf* power allows us to extend τ_{rel} period to 500 ms and beyond.

In the pulse sequence Fig. 2 proton decoupling is extended to also cover t_1 evolution period. It is well known that proton decoupling during t_1 generally leads to sharper spectral peaks [47,48]. This is particularly important for IDPs where the spectra are highly crowded. Under physiological conditions, IDPs experience fast solvent exchange which efficiently destroys two-spin modes such as $2N_xH_z$ [49]. For conventional HSQC experiment, this causes substantial line broadening in ^{15}N dimension resulting in poor-quality

spectra. In contrast, decoupled HSQC experiment records the evolution of pure in-phase coherence, N_x, and therefore avoids this problem. The sequence shown in Fig. 2 is conceptually similar to the previously published scheme [50], although the objectives of the two experiments and the technical details are different (discussed later in the article).

As discussed in detail below, the ^{15}N 180° CPMG pulses should be calibrated with care for the purpose of PD-CPMG measurements. However, standard calibration scheme is sufficient – we recommend using the HSQC-based calibration experiment.[1] The calibration of proton 90° pulse for DIPSI-2 decoupling does not need to be particularly accurate since PD-CPMG sequence is highly tolerant to pulse calibration errors. It is advisable to use the HSQC-based calibration experiment, but water-based calibration is also acceptable. It is important, however, that the duration of τ_{CP} is calculated rigorously according to the synchronization condition. For example, the experiment described in Fig. 2 was set up with $2\tau_{CP}$ = 13.9676 ms, even though for the sake of brevity we refer to it as $2\tau_{CP}$ = 14 ms (see figure caption for details).

Note that fast solvent exchange partially cancels the advantage of the sensitivity enhancement in the pulse sequence Fig. 2. A maximum potential gain from the use of the sensitivity enhancement scheme is given by a factor $\sqrt{2}$. However, the sequence requires an extra transfer step of ca. 5 ms duration. Assuming that amide solvent exchange rate is 100 s^{-1}, it is easy to estimate that the loss of signal due to solvent exchange completely offsets the sensitivity gain. Nevertheless, for most samples of practical interest the exchange rate is lower than 100 s^{-1} and therefore the use of the sensitivity enhancement scheme is likely justified (also due to its excellent water suppression properties).

Finally, under the conditions of fast solvent exchange there is a potential risk associated with partial saturation of the water signal [57,58]. Generally speaking, the degree of saturation is variable and, in particular, depends on τ_{rel}. Since the saturation is effectively transmitted to the amide sites, the magnitude of $^1H^N$ magnetization available prior to each scan also turns out to be dependent on τ_{rel}. This causes an unwanted modulation of the measured relaxation curves, which can seriously compromise the accuracy of the experiment [59,60]. To avoid these artefacts we have chosen to "crush" water magnetization following the acquisition period and prior to the recycling delay, thus creating highly reproducible initial conditions. This design leads to a very accurate experiment. In principle, the sensitivity is to a certain degree sacrificed relative to a sequence that preserves water magnetization. However, in practice the sensitivity is similar to the standard ^{15}N relaxation experiments where the water signal is strongly saturated with no regard for solvent exchange.

2.2. The results from PD-CPMG are consistent with ^{15}N spin-lock experiment

In order to confirm that PD-CPMG scheme achieves consistency with other commonly used R_2 measurement strategies such as ^{15}N spin-lock experiment, we have collected both types of data on the samples of chicken α-spectrin SH3 domain (1.5 mM) and unfolded ubiquitin (2.0 mM). Only those peaks that do not have any significant spectral overlaps have been selected for processing and analyses (58 residues for α-spc SH3 domain and 48 residues for unfolded ubiquitin). The quality of the relaxation curves is excellent in both experiments, with very low fitting residual. Importantly, the results of the two independent measurements show an excellent correlation with each other, as evidenced by the correlation coefficients 0.999 and 0.993, see Fig. 3. The systematic deviation between the two datasets is very small. Specifically, the difference between the mean values of R_2 from the new PD-CPMG experiment and the standard spin-lock experiment is +0.9% in the case of α-spc SH3 and −1.3% in the case of ubiquitin. Even though such differences are negligibly small, we have a fairly good idea of what causes them. In the case of ubiquitin, the conductivity of the acidic sample is relatively high and therefore it experiences a substantial amount of heating [61]. PD-CPMG sequence heats the sample a little more than spin-lock sequence, 0.45 vs. 0.30 °C, leading to slightly lower R_2 rates. For α-spc SH3 the amount of heating generated by both experiments is small, less than 0.1 °C. However, one has to bear in mind that α-spc SH3 is a folded protein with broadly dispersed $^1H^N$ spectrum. DIPSI-2 decoupling does not work quite as well in this case, leading to a small magnetization loss and consequently slightly overestimated R_2 rates. Both aspects – the performance of proton composite decoupling and sample heating – are discussed in greater detail in what follows.

We have also recorded a control PD-CPMG experiment where the strength of the DIPSI-2 decoupling was lowered from ca. 4 kHz to ca. 3 kHz. The results remain virtually unchanged, as illustrated in Fig. 4 (correlation coefficients 0.999 and 0.998). The observed amount of systematic deviation is vanishingly small (and can be rationalized along the same lines as above). Importantly, this result means that PD-CPMG experiment with relatively low level of decoupling power, ca. 3 kHz, can be used to sample the relaxation profiles up to $\tau_{rel} \sim 1$ s and beyond. It can also be used to collect the data from samples with high conductivity (high ionic strength) where sample heating due to deposition of the *rf* power needs to be reduced.

2.3. PD-CPMG scheme improves the precision of ^{15}N R_2 measurements

The PD-CPMG sequence is intended to improve the sampling of ^{15}N relaxation curves in the situation when R_2 relaxation is relatively slow, on the order of 1–3 s^{-1}. Specifically, the experiment is relevant for IDPs at or near the physiological temperature. Of necessity, NMR studies of IDPs are often conducted at low concentration, 100 μM or less. Furthermore, the time frame of the measurements is usually limited, since the samples have a tendency to rapidly deteriorate. For those samples where pH is close to physiological, the signal-to-noise ratio is further lowered by amide solvent exchange which in the case of IDPs affects all residues in the protein. Under these circumstances, the precision of ^{15}N R_2 data can become a critical issue.

In order to model this type of a situation, we have prepared two samples of unfolded ubiquitin with protein concentration 100 and 50 μM and collected the relaxation data at 37 °C. Toward this end, we employed the standard ^{15}N $R_{1\rho}$ experiment [10] as well as the new PD-CPMG scheme. All experiments were recorded in duplicate to assess the reproducibility of the results. The outcome is illustrated in Figs. 5 and 6.

The relaxation curves recorded in the repeat spin-lock and PD-CPMG experiments are illustrated in Fig. 5. The results for residue L15 are shown as is, i.e. in a form of $R_{1\rho}$ decay and apparent R_2^{app} decay, with no correction for ^{15}N frequency offset. The quality of the fits is similar for the two experiments, as can be appreciated from the magnitude of the fitting residual χ. However, the reproducibility of the relaxation rates as determined in the spin-lock measurements (2.41 and 2.81 s^{-1}) is clearly worse than for PD-CPMG (2.67 and 2.56 s^{-1}). This happens because the spin-lock experiment is limited to short τ_{rel} delays and therefore cannot properly sample the relaxation decay profile [34].

The summary of the data from the repeat experiments is shown in Fig. 6. Clearly, PD-CPMG achieves a much better reproducibility

[1] Specifically, 1D version of $^1H^N$–^{15}N HSQC, where a variable-length nitrogen pulse is inserted following the initial INEPT period (Daiwen Yang and Lewis Kay, unpublished).

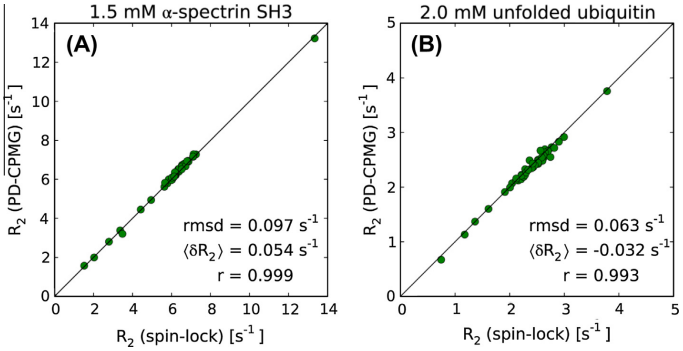


Fig. 3. Comparison of ^{15}N R_2 rates from the PD-CPMG experiment and the standard spin-lock measurement [10]. Each experiment took ca. 2 h to record. The PD-CPMG experiment used 4 kHz DIPSI-2 decoupling. All datasets have been corrected for ^{15}N frequency offset effects (see below for details). PD-CPMG experiment on the sample of α-spc SH3 was carried out with $2\tau_{CP}$ = 13.7373 ms and the set of six τ_{rel} delays ranging from 0 to 275 ms.

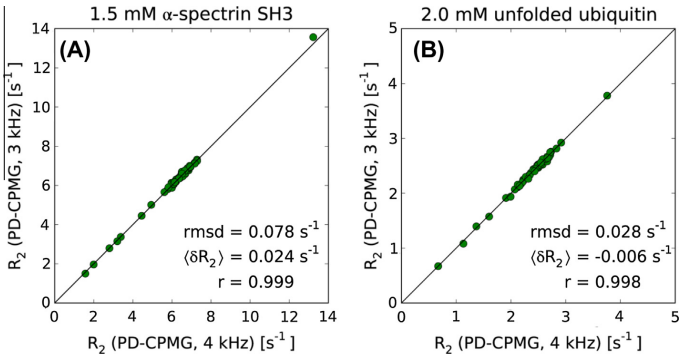


Fig. 4. Comparison of ^{15}N R_2 rates from the PD-CPMG measurements using ca. 4 kHz and 3 kHz DIPSI-2 decoupling. In the experiments with 3 kHz decoupling, the data from α-spc SH3 were obtained using m = 1, $2\tau_{CP}$ = 9.6509 ms, and the set of six τ_{rel} delays ranging from 0 to 193 ms, whereas the data from ubiquitin were collected using m = 1, $2\tau_{CP}$ = 9.8235 ms and the set of six τ_{rel} delays ranging from 0 to 589 ms.

than the standard spin-lock measurement [10]. The average absolute-value error in $R_{1\rho}$ data is three times as large as in PD-CPMG R_2^{app} data: 9.1% vs. 3.7% for 100 μM sample and 22.6% vs. 7.5% for 50 μM sample. This difference in precision can be critical for studies that use ^{15}N relaxation to quantify changes in IDPs conformational dynamics in response to various perturbations, e.g. point mutations, posttranslational modifications, variations in pH, crowding effects, etc. Such changes are typically subtle and can be successfully identified only when high-precision ^{15}N relaxation data are available. Note that $R_{1\rho}$ data shown in Fig. 6c are essentially useless in this regard since the measurement errors are of the same magnitude as per-residue R_2 variations. In contrast, R_2^{app} data shown in Fig. 6d retain the ability to distinguish between mobile residues, constrained residues, and those residues that experience exchange broadening.

2.4. Modified PD-CPMG scheme for arginine side-chain $^{15}N^{\varepsilon}$

After small modifications, the PD-CPMG scheme Fig. 2 can also be used to measure R_2 relaxation rates of $^{15}N^{\varepsilon}$ spins in Arg side chains. Arginine side chains are long and flexible; in globular proteins they are typically immersed in solvent and move in a relatively unconstrained fashion. Consequently, $^{15}N^{\varepsilon}$ R_2 rates tend to be low compared to the backbone ^{15}N rates and thus particularly well suited for PD-CPMG measurements. It is also worth noting that N^{ε} are more prone to solvent exchange than the backbone amide sites. Therefore, the favorable properties of the PD-CPMG experiment with regard to solvent exchange are especially useful in the context of $^{15}N^{\varepsilon}$ measurements.

The pulse sequence requires some alterations, as described in the caption of Fig. 2. The main source of difficulties is the weak $^{15}N^{\varepsilon}$–$^{15}N^{\eta}$ two-bond coupling with the magnitude ca. 1 Hz [62]. This coupling remains active during the standard CPMG pulse train which employs hard nitrogen pulses. Since τ_{rel} times employed in our measurements are long, the $^2J_{N\varepsilon N\eta}$ coupling achieves partial conversion of N_x^{ε} into $2N_y^{\varepsilon}N_z^{\eta}$ during this time period. This causes partial loss of signal and distorts the observed relaxation decay profiles.

Since $^{15}N^{\varepsilon}$ and $^{15}N^{\eta}$ have significantly different chemical shifts (separated in the spectrum by ca. 10 ppm), this problem can be easily overcome by application of selective 180° pulses during the CPMG period. For this purpose we have employed REBURP

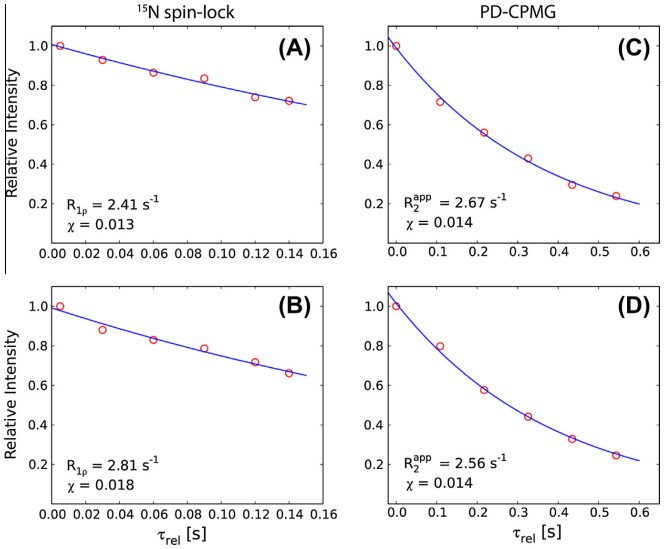


Fig. 5. Relaxation decay profiles for residue L15 in 0.1 mM sample of unfolded ubiquitin as obtained from the repeat measurements using ^{15}N spin-lock experiment (panels A and B) and PD-CPMG (panels C and D). For reference, the measurements using 2.0 mM sample of unfolded ubiquitin yield $R_{1\rho} = R_2^{app} = 2.58\ s^{-1}$ (uncorrected for ^{15}N offset effects). Parameter χ indicated in the plot is the rmsd between the experimental and fitted intensities.

pulses which have been used in the context of CPMG measurements before [63,64]. One of the potential problems with such strategy is that REBURP pulses are relatively long, 4.0 ms, so that nitrogen magnetization spends part of the time away from the transverse plane. As a result, the measured relaxation decay may potentially represent a mixture of R_2 and R_1 rates. However, numerical simulations suggest that this effect introduces only a very small error, on the order of 1%. To understand this result, one should take into consideration several factors. (i) REBURP pulses are applied relatively infrequently, $2\tau_{CP} \geqslant 10$ ms. (ii) REBURP pulses are applied on-resonance or close to resonance (since $^{15}N^{\varepsilon}$ chemical shifts are confined to a relatively narrow range). Under these conditions the magnetization mostly remains in the transverse plane during the REBURP pulse. (iii) REBURP pulses are applied with [0013] phase cycle. Considering N_x magnetization, we observe that it is locked by the first two pulses, such that no R_1 character is introduced into spin relaxation. Similarly N_y magnetization is locked by the last two pulses. (iv) In rapidly moving Arg side chains, R_2 and R_1 rates tend to be similar.

Ultimately we rely on the comparison with $R_{1\rho}$ experiment to confirm that PD-CPMG measurement is error-free. Importantly, $R_{1\rho}$ experiment is insensitive to the effect of $^2J_{N\varepsilon N\eta}$. Numerical simulations show that given a relatively large chemical shift offset between $^{15}N^{\varepsilon}$ and $^{15}N^{\eta}$, ca. 10 ppm, moderate spin-lock strength, ~1.5–2 kHz, very weak coupling, ~1 Hz, and relatively short τ_{rel} delays, ~100 ms, it is safe to neglect Hartmann–Hahn transfer during the spin-lock period. In Fig. 7 we compare the R_2 rates obtained from the standard $R_{1\rho}$ experiment with those measured by $^{15}N^{\varepsilon}$ PD-CPMG. The two datasets are consistent within 0.5%, which is approximately the same as the level of agreement previously observed for the backbone amide sites, see Fig. 3a. This result provides validation for the modified version of the PD-CPMG sequence aimed at $^{15}N^{\varepsilon}$ spins. Potential gains in precision in this case should be even more significant than for backbone measurements.

3. Experimental details

3.1. ^{15}N offset dependence

One significant shortcoming of the traditional CPMG experiment is that it requires a numerical procedure to correct for ^{15}N offset effect [65]. In contrast, the PD-CPMG experiment using [0013] phase cycle allows for a simple analytical correction [44]:

$$R_2^{app} \approx R_2 + (R_1 - R_2)\mu^2 \tag{1}$$

$$\mu = \nu_{off}^{N}/\nu_1^{N(180)}$$

Here R_2^{app} is the apparent decay rate as derived directly from the PD-CPMG measurement, ν_{off}^{N} is the nitrogen frequency offset for a given amide resonance, and $\nu_1^{N(180)}$ is the field strength of the hard ^{15}N 180° CPMG pulses in the units of Hz. For the experimental settings used in this study the magnitude of μ^2 does not exceed 0.03. Consequently, it is safe to ignore higher-order corrections which are proportional to μ^4.

The formula Eq. (1) reflects the fact that during τ_{rel} period (or more specifically during $2\tau_{CP}$ intervals) ^{15}N magnetization moves on a path which takes it away from the transverse plane. By the end of each [0013] cycle, however, the magnetization is returned to the transverse plane as a result of the good self-compensation properties of this cycle. The original correction formula published by Bain et al. contains an extra term which describes spin relaxation during ^{15}N 180° pulses [44]. In our case this additional term can be safely neglected because ^{15}N 180° pulses are very short compared to $2\tau_{CP} \geqslant 10$ ms. The resulting simple expression Eq. (1) can be easily inverted to determine the true relaxation rate constant R_2.

The validity of Eq. (1) has been tested by means of numeric simulations. Briefly, spin evolution in the two-spin ($^1H^N$, ^{15}N) system was modeled using full product operator basis of dimension 15.

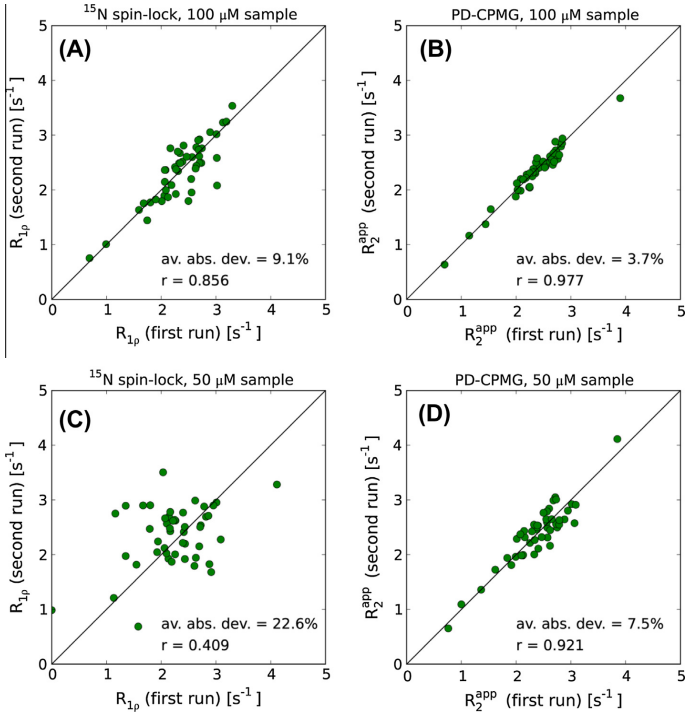


Fig. 6. The reproducibility of $R_{1\rho}$ and R_2^{app} relaxation rates as measured in the repeat spin-lock and PD-CPMG experiments using low-concentration samples of unfolded ubiquitin at 37 °C. Each experiment was recorded in a total of ca. 2 h.

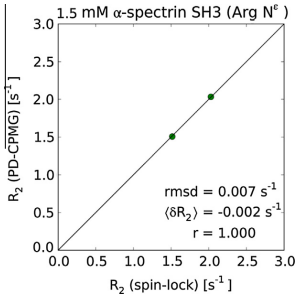


Fig. 7. Comparison of arginine $^{15}N^{\varepsilon}$ R_2 rates from the PD-CPMG experiment and the standard spin-lock measurement. The results are from two arginine residues in α-spc SH3 domain, R21 and R49. Protein concentration in the sample is 1.5 mM. The PD-CPMG sequence has been adapted to measure $^{15}N^{\varepsilon}$ relaxation as discussed in the text. In running this experiment one has to be mindful of $^{15}N^{\varepsilon}$ and $^{15}N^{\eta 1}$ chemical shifts specific to a given protein. The spin-lock sequence [10] was used as is, with nitrogen *rf* carrier set to 85 ppm. Both datasets have been corrected for ^{15}N frequency offset effects.

All coherent evolution mechanisms were included (chemical shift, *rf* pulses, scalar coupling) as well as the standard relaxation mechanisms (dipolar and CSA). In terms of *rf* pulses, the simulations reproduce the experimental setup described in the caption of Fig. 2. We have also included the effects of ^{15}N *rf* field inhomogeneity which was modeled via Gaussian distribution with the width $2\sigma = 0.1 \times \nu_1^N$. Further details can be found in Materials & Methods.

In Fig. 8a we illustrate the dependence of the apparent R_2^{app} rate on ^{15}N frequency offset ν_{off}^N. Shown are the simulated R_2^{app} rates (red circles) as well as the predictions using Eq. (1) (black curve). The true value of R_2 is indicated by the horizontal dashed line. This result suggests that the rates measured via PD-CPMG experiment and corrected by means of Eq. (1) provide a very accurate measure of R_2 relaxation.

Note also that the offset dependence illustrated in Fig. 8a is rather weak. This is not surprising since PD-CPMG employs a train of hard ^{15}N 180° pulses which effectively refocus nitrogen magnetization across the entire spectrum. In fact, if the offset dependence is ignored the resulting error in determination of R_2 does not exceed 1%. In most cases the error of this magnitude can be tolerated and therefore it is safe to assume that $R_2 = R_2^{app}$. This obviates the need in ^{15}N R_1 measurement which is otherwise needed to calculate the offset correction according to Eq. (1). Finally, it is worth noting that the result shown in Fig. 8a is independent of the *rf* field

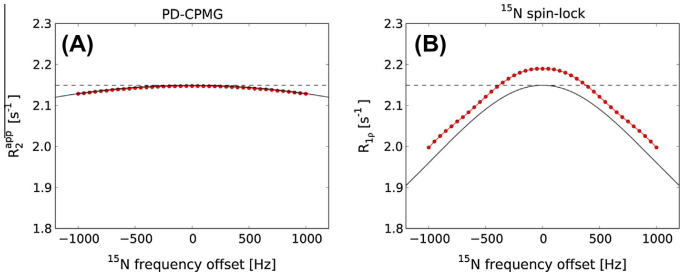


Fig. 8. Simulated dependencies (A) R_2^{app} vs. ν_{off}^N for the PD-CPMG experiment and (B) $R_{1\rho}$ vs. ν_{off}^N for the spin-lock experiment. The results from numeric simulations are shown with red circles (red lines), the calculations using analytical formulas are shown with solid black lines, the target value of R_2 is shown with the horizontal dashed line. The CPMG and spin-lock pulse sequence elements are modeled with the same settings as used in the experimental study (see Fig. 2 and Section 5). In particular, proton DIPSI-2 decoupling is assumed to be applied on-resonance with the *rf* field strength 4 kHz.

inhomogeneity on both ^{15}N and ^{1}H channels (assuming any reasonable *rf* field distribution).

For comparison, in Fig. 8b we show the equivalent dependence for the spin-lock experiment, $R_{1\rho}$ vs. ν_{off}^N [10]. Notably, there is a shift between the simulated $R_{1\rho}$ rates (red circles) and the standard correction formula, $R_{1\rho} = R_2 \sin^2\theta + R_1 \cos^2\theta$ (black curve). The shift, which has a magnitude of 0.03–0.04 s^{-1}, can also be found in the Korzhnev's original report (see Figs. 2 and 3 therein). It arises from the interplay between the spin-lock, $^{1}J_{NH}$ coupling, and ^{1}H 180° pulses used to suppress dipolar-CSA cross-correlations [10]. Even though such pulses are applied infrequently, they can still give rise to a discernible error. In relative terms, this error can be safely neglected for slowly tumbling globular proteins, but not necessarily so for highly mobile IDPs.

As it turns out, in the case of high-conductivity ubiquitin sample the systematic error associated with spin-lock experiment, Fig. 8b, is compensated by the effect from sample heating (see below). Such error cancelation, however, is fortuitous. For samples with low conductivity the error would persist at the level of ca. 2%. Furthermore, the magnitude of the error is sensitive to the details of ^{15}N *rf* field inhomogeneity and may potentially increase beyond this level.

3.2. ^{1}H decoupling and choice of τ_{CP}

As already pointed out, proton decoupling should be synchronized with ^{15}N CPMG pulse train according to

$$2\tau_{CP} + p_{N(180)} = m\Delta_{DIPSI2} \quad (2)$$

For this purpose, one should calibrate the proton 90° pulse at the chosen decoupling power (3–4 kHz), calculate the duration of the decoupling cycle Δ_{DIPSI2}, and then set τ_{CP} accordingly (see caption of Fig. 2).

What happens if the synchronization condition is not met? DIPSI-2 sequence is designed to achieve near-complete suppression of $^{1}J_{NH}$ over the time period Δ_{DIPSI2}. However, if at some point during Δ_{DIPSI2} one applies a ^{15}N 180° pulse this would interfere with the decoupling scheme, leading to less-than-perfect suppression of $^{1}J_{NH}$. As a consequence, nitrogen spin would experience a certain amount of scalar-coupled evolution and suffer from additional magnetization loss (e.g. due to dephasing of $2N_yH_z$ under the effect of the proton *rf* field). From the perspective of PD-CPMG measurements, this would result in overestimated nitrogen R_2 rates. Indeed, we have been able to reproduce this effect via simulations and experimentally. When PD-CPMG experiment is set up with no regard for synchronization conditions, the determined R_2 rates prove to be systematically higher. For example, in trial experiments involving unfolded ubiquitin we observed the bias of ca. 0.07 s^{-1}. While such relatively small deviations can usually be neglected for slowly tumbling globular proteins, they are not necessarily insignificant in the case of IDPs.

The undesirable effects can be avoided by enforcing the synchronization condition Eq. (2). In principle, the value of m used in the synchronization condition should be even. However in practice odd values of m are equally acceptable (verified both experimentally and by simulations). This is so because one-half of the standard DIPSI-2 cycle still offers excellent decoupling properties. For example, in the case of ubiquitin measurements described in the caption of Fig. 2 the synchronization condition reads $2\tau_{CP} = (m \cdot 7.0218 - 0.0760)$ ms, where m = 1, 2, 3, ...

The choice m = 2 corresponds to the strength of the effective *rf* field $\nu_1^{CPMG} = 1/4\tau_{CP} = 36$ Hz. This is much lower than the strength of the *rf* field used for proton decoupling, 4.1 kHz. Thus, the experiment should be safe with regard to a potential risk of Hartmann–Hahn transfer. Indeed, our simulations predict that the outcome of the experiment is independent of the choice of m, confirming that Hartmann–Hahn transfer is not an issue.

However, the experimental tests revealed a different type of complication. As it turns out, the measurements using large values of m (i.e. long τ_{CP} delays) yield slightly higher than expected R_2 rates. For instance, in the case of ubiquitin, increasing m from 2 to 6 leads to 0.1 s^{-1} increase in the average R_2 value. The trend is reverse of what could be expected from Hartmann–Hahn cross-polarization. It also cannot be explained by a variable degree of sample heating (the effect is also observed in α-spectrin SH3, where there is very little heating). We have considered a number of possible scenarios: (i) dispersion-like effect associated with pervasive small R_{ex} contributions; (ii) the effect from small scalar couplings between ^{15}N and $^{1}H^{\alpha}$, $^{1}H^{\beta}$; (iii) the effect from small scalar couplings between ^{15}N and natural-abundance ^{13}C; (iv) the effect of dipole–dipolar cross-correlations in a three-spin system ^{15}N, $^{1}H^{N}$, and $^{1}H^{\alpha}$ ($^{1}H^{\beta}$); (v) the effect of diffusion (convection flow) in inhomogeneous static magnetic field. Using additional experimental tests and/or numeric simulations we were able to rule out each of these explanations.

Instead, we hypothesize that the performance of DIPSI-2 decoupling is not quite as efficient as suggested by numeric simulations. Small instrumental defects such as finite pulse rise time, phase transients, or amplifier drooping have a subtle, yet discernible effect on the real-life performance of the decoupling sequence. To demonstrate the feasibility of this hypothesis we conducted additional experimental measurements. Specifically, the PD-CPMG experiment was repeated with ^{1}H *rf* carrier shifted from 8.3 ppm to 6.5 ppm. The numeric simulations (including carefully designed

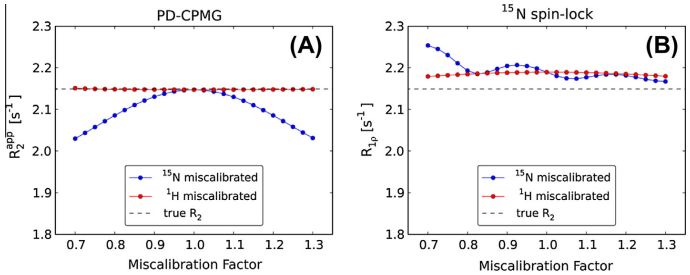


Fig. 9. The effect of *rf* pulse miscalibration on (A) R_2^{app} rates from PD-CPMG experiment and (B) $R_{1\rho}$ rates from spin-lock experiment. The miscalibration factor represents the ratio of the actual ν_1 to the target ν_1. The simulations have been conducted assuming $\nu_{off}^N = 0$ and $2\sigma = 0.1 \times \nu_1^N$. For proton channel, the effects of frequency offset and *rf* field inhomogeneity were found to be negligibly small.

three-spin simulations) predict that such carrier jump should have no effect on the measured R_2 rates. Yet the experimentally measured R_2 values in the sample of unfolded ubiquitin increased on average by 0.03 s^{-1}. In the case of α-spc SH3 with its widely dispersed $^1H^N$ spectrum the increase was more substantial, 0.06 s^{-1}. This lends support to our hypothesis that the real-life performance of proton decoupling is less perfect than can be inferred from the simulations.

This result makes us suggest that DIPSI-2 decoupling is less efficient than expected in suppressing $^1J_{NH}$ coupling. In this situation we rely on ^{15}N 180° pulses to assist with decoupling. In order to fully suppress $^1J_{NH}$ evolution, nitrogen pulses should be applied sufficiently frequently. It is also important that nitrogen pulses are applied with maximum power, in our case 6.6 kHz.[2] To summarize, the PD-CPMG experiment presented in this paper should be executed with proton decoupling *rf* power of 3–4 kHz and $2\tau_{CP}$ - ~ 10–15 ms (subject to synchronization condition, m = 1–2). Use of the lower decoupling power or longer τ_{CP} delays results in a small but measurable bias, leading to overestimated R_2 rates.

Finally, we would like to comment on the choice of proton decoupling scheme. DIPSI-2 was selected primarily for its excellent J-coupling suppression properties. The decoupling bandwidth is a secondary consideration (particularly for disordered proteins with their narrow range of $^1H^N$ chemical shifts). The PD-CPMG measurements using DIPSI-3 decoupling produce the results that are virtually identical to those obtained with DIPSI-2. Choosing between the two, we prefer the latter scheme since it offers a benefit of a shorter cycle. Accurate results have also been obtained using WALTZ-16. However, in this case the experiment becomes more sensitive to synchronization condition – away from the correct setting expressed by Eq. (2) the measured R_2 rates turn out to be significantly overestimated (confirmed both experimentally and by simulations). Along these lines we can explain the appearance of bias in the earlier implementations of PD-CPMG [38]. Finally, GARP-1 decoupling sequence shows relatively poor performance. This is not surprising since GARP offers excellent bandwidth, but mediocre J-coupling suppression properties [66].

3.3. Rf pulse miscalibration

The numeric simulations modeling the effect of *rf* pulse miscalibration on the measured R_2^{app} and $R_{1\rho}$ rates are illustrated in Fig. 9. The PD-CPMG experiment turns out to be remarkably stable with respect to miscalibration of proton pulses in the DIPSI-2 decoupling scheme (red circles[3] in Fig. 9a). This reflects excellent self-compensation properties of the DIPSI-2 sequence as manifested in our simulations. On the other hand, the miscalibration of ^{15}N 180° pulses leads to nitrogen magnetization spending part of the time away from the transverse plane. This results in the increased R_1 character of the spin relaxation and lower R_2^{app} rates (blue circles). Note that in practice ^{15}N pulse calibration should be accurate to at least within ±5%. Hence the actual miscalibration factor is likely confined to the narrow range 0.95–1.05. In this range the R_2^{app} rates obtained from the PD-CPMG measurements remain essentially unchanged. We have also considered the scenario where both ^{15}N and ^{1}H pulses are miscalibrated. The results are no different from the situation where ^{15}N pulses alone are misset.

As a point of comparison, the effect of *rf* field miscalibration on $R_{1\rho}$ experiment is also illustrated, Fig. 9b. There is certain amount of error associated with this experiment as manifested in the shift between the simulated and the target relaxation rates (see Fig. 8 and related discussion). Varying the strength of on-resonance ^{15}N spin lock in the interval from 1.2 to 2.2 kHz modulates the magnitude of error as shown in the plot; the sinusoidal modulation pattern is similar to the one previously observed [10]. There is also an additional element of bias associated with spin-lock experiment which is not reflected in Fig. 9b. The failure to correctly determine ν_1^N leads to an error in applying the ^{15}N offset correction. The amount of error can be substantial, on the order of 0.1 s^{-1}. This is much worse than in the case of PD-CPMG experiment, which is fairly insensitive to the ^{15}N offset effects (see Fig. 8).[4]

Finally, we have conducted additional experimental measurements to test the validity of the simulations illustrated in Fig. 9a. For this purpose we have deliberately misset the proton pulse in the DIPSI-2 decoupling sequence – instead of the properly calibrated value, $p_{H(90)}$ = 61 μs, we have used 56 or 66 μs. Using this altered experimental setup, we repeated the PD-CPMG measurements on the sample of unfolded ubiquitin. The results proved to be in perfect agreement with the previously obtained data, thus confirming that PD-CPMG experiment is immune to miscalibration of the DIPSI-2 proton pulses. Along the same lines we have tested the incorrect settings for nitrogen CPMG pulses: instead of $p_{N(180)}$ = 76 μs we have used 66 or 86 μs. As expected, this led to appreciable decreases in the measured R_2 rates (on average by 0.09 and 0.06 s^{-1}, respectively). One should keep in mind, however, that these tests involve a significant element of exaggeration. In

[2] If CPMG pulses are applied with reduced level of power (4 kHz) some of the recorded decay curves become slightly distorted. This effect occurs only at certain specific values of ^{15}N frequency offset.

[3] For interpretation of color in Fig. 9, the reader is referred to the web version of this article.

[4] As for the proton pulses, their sole purpose in the spin-lock experiment is to suppress the dipolar-CSA cross-correlation. In this role they remain efficient even when their actual flip angle is varied between 130° and 230° (red circles in Fig. 9b).

practice, hard ^{15}N pulses are calibrated with much better accuracy, on the order of ±1 μs. Under these conditions the PD-CPMG experiment is essentially insensitive to the effects of pulse miscalibration.

3.4. The importance of using external ^2H lock

It is clearly desirable to study IDPs at or near physiological conditions. However, so far such studies have been rare [67]. The main difficulties arise from fast solvent exchange, $k_{ex} \sim 10$–$100\ s^{-1}$, which undercuts the sensitivity of HSQC experiment and worsens the situation with spectral overlaps. There are also other, more technical, problems associated with solvent exchange that can compromise the accuracy of ^{15}N relaxation measurements.

Most biological NMR samples are prepared with 5–10% D_2O for the purpose of frequency locking. However, under fast solvent exchange conditions the presence of D_2O in the sample could have a number of undesirable consequences for ^{15}N R_2 measurements [12]. (i) Standard relaxation experiments use $^1H^N(^{15}N)$ sites as a point of origin. During the course of the sequence, and in particular during τ_{rel}, the original $^1H^N(^{15}N)$ groups are partially converted into unobservable $^2H^N(^{15}N)$. This transition process modulates the recorded decay profile and thus interferes with R_2 determination. (ii) If solvent exchange is fast on the time scale of τ_{rel}, the measured R_2 rate represents a weighted average of $^{15}N(^1H^N)$ and $^{15}N(^2H^N)$ relaxation rates. (iii) In the latter case, $^{15}N(^2H^N)$ relaxation contains a sizable contribution from the scalar mechanism due to fast deuterium relaxation and solvent exchange. In PD-CPMG experiment this contribution is only partially suppressed by infrequent application of 180° ^{15}N pulses. (iv) In addition, the apparent R_2 may include an R_{ex}-type term which stems from the isotopic frequency shift between $^{15}N(^1H^N)$ and $^{15}N(^2H^N)$.

To avoid all of these complications, we use the external frequency lock, i.e. 100 μL D_2O placed in a coaxial insert fitted into the standard NMR tube. Such inserts can be purchased inexpensively from a number of vendors (we used WGS-5BL from Wilmad). They have been widely used in the past in the studies dealing with fast solvent exchange [68,69]. It has been shown that the samples using coaxial inserts are less likely to suffer from convection flow effects [70]. On the downside, such inserts are currently unavailable for Shigemi tubes.

3.5. Manipulating water magnetization

As discussed above, PD-CPMG sequence is intended for use with intrinsically disordered proteins at physiological temperature. If sample pH is also close to physiological, then solvent exchange becomes a major factor. Specifically the accuracy of the measurements can be compromised by the transfer of the saturation from water to amides mediated by fast exchange. Other transfer mechanisms, such as (i) intermolecular NOE and (ii) solvent exchange involving hydroxyl and amine groups followed by intramolecular NOE, can potentially add to the problem. Additional damage can be caused by radiation damping.

There are several ways of dealing with this complication. For instance, one can try to preserve water magnetization during the course of the pulse sequence. In principle, this is feasible since DIPSI-2 sequence can be used to lock water magnetization [71]. However, in practice this method is not well suited for our experiment. During the long τ_{rel} delays, ~ 0.5–1 s, spin-locked water magnetization undergoes significant relaxation decay and therefore fails to be conserved. The alternative approach is to use very long recycling delays allowing for complete recovery of water magnetization. For obvious reasons this is also impractical.

Instead, we have chosen to apply water purge element [51] following the acquisition period and prior to the recycling delay. This method guarantees that the amount of water magnetization present at the beginning of each scan always remains the same. Consider the situation where solvent exchange is fast on the time scale of spin–lattice relaxation, $k_{ex} \sim 10\ s^{-1}$ or higher. In this case the recovery of amide magnetization is controlled by the water R_1 rate, which is $0.2\ s^{-1}$ at 37 °C [72]. Under these circumstances, the optimal recycling delay is approximately 5.0 s. Long recovery time means that the sensitivity of the experiment suffers – it is ca. 2-fold lower compared to the schemes that preserve water magnetization. However, this is the sacrifice that is necessary to ensure the accuracy of R_2 measurements. Note that there is also a positive aspect to longer recycling delays as they help to reduce sample heating (discussed below).

The same problems are encountered in the standard relaxation experiments when applied to IDPs at or near physiological conditions. For example, in the standard $R_{1\rho}$ experiment the status of water magnetization changes depending on the number of ^1H 180° pulses applied during τ_{rel}. This gives rise to a distinctive modulation pattern in the recorded relaxation curves (data not shown). To avoid such spurious effects, we have also inserted the water-crusher element in the standard $R_{1\rho}$ pulse sequence [10]. All $R_{1\rho}$ measurements reported in this work were conducted using this amended version of the original sequence (see Materials & Methods).

3.6. Contribution from solvent exchange to R_2^{app}

Above we have discussed two potential complications associated with solvent exchange: (i) proton-deuterium exchange in a D_2O-containing sample and (ii) saturation transfer from H_2O to amide protons. In this section we discuss a third mechanism whereupon solvent exchange can interfere with ^{15}N R_2 measurements. In the absence of proton decoupling, the in-phase N_x magnetization interconverts with the anti-phase $2N_yH_z$. The latter is destroyed by solvent exchange, which physically separates the two correlated spins [49]. When solvent exchange is sufficiently fast on the time scale of $^1J_{HN}$ the resulting loss of magnetization can be described as scalar relaxation of the second kind.

In the PD-CPMG and spin-lock experiments, $^1J_{HN}$ evolution during τ_{rel} is suppressed. Therefore, one may expect that scalar relaxation should be absent. This is not quite so. For spin-lock experiment, the standard expression for scalar relaxation rate $R_2^{sc} = (\pi J_{HN})^2 \tau_{ex}/(1 + (2\pi\nu_1\tau_{ex})^2)$ [73] immediately suggests that the effect can be significant.[5] A similar outcome can be expected for the PD-CPMG experiment, although no analytical formula is available to address the situation involving DIPSI-2 decoupling.

The simulated $R_{1\rho}$ and R_2^{app} constants for a range of solvent exchange rates k_{ex} are shown in Fig. 10. The graph demonstrates that slow exchange, $k_{ex} < 10\ s^{-1}$, has virtually no effect on the extracted relaxation rates. As the exchange rate increases to 10–$100\ s^{-1}$, a significant amount of error is generated. The $R_{1\rho}$ experiment using 1 kHz spin lock is particularly vulnerable (green curve). The $R_{1\rho}$ experiment using 2 kHz spin lock and PD-CPMG experiment using 4 kHz DIPSI-2 decoupling produce a similar amount of error, which is not necessarily negligible (orange and red curves, respectively). One should be mindful of this effect when conducting measurements under somewhat extreme conditions where k_{ex} exceeds several tens of s^{-1}. In this situation it may be advisable to collect the data using higher decoupling power, which helps to eliminate the contribution from solvent exchange into R_2^{app}. For example, we have successfully recorded PD-CPMG experiment with 6 kHz DIPSI-2 decoupling and τ_{rel} extending to 640 ms (data not shown). Under certain circumstances one may also want to collect the data at

[5] In this expression $\tau_{ex} = k_{ex}^{-1}$ and ν_1 is the strength of the spin-lock.

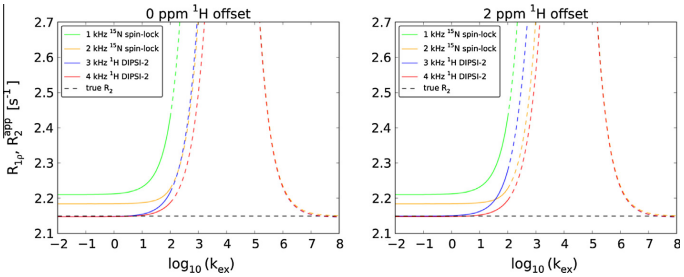


Fig. 10. Dependence of $R_{1\rho}$ and R_2^{app} rates from ^{15}N spin-lock and PD-CPMG experiments on solvent exchange rate k_{ex}. The dependence arises from the loss of $^1H^N$–^{15}N spin correlations due to amide solvent exchange. The simulations have been conducted using the same model as in Fig. 8, with $\nu_{off}^N = 0$. Solid lines correspond to the conditions of practical interest; black dashed horizontal line represents the true value of R_2. k_{ex} is given in the units of s^{-1}.

two decoupling field strengths, 3 and 4 kHz, to identify the rates affected by solvent exchange.

The range $k_{ex} \sim 10–100\ s^{-1}$ is relevant for IDPs under physiological conditions [74]. This is also the range where ^{15}N relaxation can be measured using the standard HSQC-based schemes. The remaining portion of Fig. 10, corresponding to ultra-fast solvent exchange, is mainly of academic interest (although suitable experimental schemes have also been developed [75,76]). The scalar contribution into $R_{1\rho}$ and R_2^{app} reaches several s^{-1} when exchange rate increases to ca. $10{,}000\ s^{-1}$ (the simulated data fall outside the area of the plot). With further increase in k_{ex} the scalar relaxation declines. This phenomenon is known as self-decoupling; in principle, self-decoupling leads to complete suppression of $^1J_{NH}$ evolution and allows for extraction of accurate R_2 rates (all profiles in the plot converge to the dashed horizontal line). For spin-lock experiment, the numeric results shown in Fig. 10 are in quantitative agreement with the calculations using Redfield-theory formula for scalar relaxation [73].

3.7. Sample heating

Sample heating is a potential concern in all heteronuclear spin–spin relaxation experiments as they deposit a copious amount of *rf* power into the sample. This is also true for the PD-CPMG scheme where proton decoupling is applied for a lengthy period of time, up to 0.5–1.0 s. The problem is exacerbated for samples with high conductivity. When using a sample of ubiquitin with pH 2.0 and recording PD-CPMG experiment with 6 kHz DIPSI-2 decoupling, we have observed heating-related artefacts such as characteristically distorted peak shapes [77]. This is attributable to the high content of H^+ ions in acidic solution; the problem disappears in the pH 3.0 sample (see below).

To mitigate the effect of heating, we have programmed the pulse sequence Fig. 2 in an interleaved fashion. In our scheme t_1 incrementation is controlled by an outer loop, whereas the array of τ_{rel} is parsed by an inner loop, alternating between short and long delays (e.g. 0, 560, 112, 448, 224, and 336 ms in the case of ubiquitin measurements). In this manner the degree of heating remains near-constant throughout the experiment. If necessary, the interleaved scheme can be further improved as described by Orekhov and co-workers [78]. The heat-equalizing element [77] can also be included into the sequence.

With the experimental setup used in our study it is straightforward to monitor the change in temperature during the PD-CPMG experiment. Toward this goal we have prepared a buffer solution of 8 M urea in D_2O, pH 3.0, which imitates ubiquitin sample. This solution was placed into NMR tube while the coaxial insert was filled with methanol. The original pulse sequence was then modified such as to allow for observation of the methanol ^{1}H spectrum (see Materials & Methods for details). This sequence was used to record a dataset which was equivalent to the original PD-CPMG in every way, except that each FID contained a strong methanol signal. The FIDs pertaining to the same τ_{rel} value have been added together, and the resulting methanol spectrum was used to determine the effective temperature associated with this particular τ_{rel} [79]. The results of this procedure are presented in Fig. 11.

Fig. 11 demonstrates that the sample heating in PD-CPMG experiment is close to uniform, as desired. In the absence of any *rf* pulsing, the temperature of the sample is determined to be 37.0 °C. The application of PD-CPMG sequence with 4 kHz DIPSI-2 decoupling and 2 s recycling delay raises the temperature to between 37.4 and 37.5 °C (green circles in the plot). When proton decoupling power is lowered to 3 kHz, the temperature is stabilized just under 37.4 °C (red circles). This is comparable to the spin-lock experiment employing τ_{rel} from 5 to 140 ms, where the similar interleaved setup heats the sample to slightly more than 37.3 °C (blue circles).

How significant is this level of heating? The data on ^{15}N R_2 relaxation in denatured ubiquitin are available at several different temperatures ([80,81], this work). From these data one can readily estimate that the temperature dependence of nitrogen R_2 is ca. $0.1\ s^{-1}/°C$. For the experiment at hand the increase in temperature amounts to ca. 0.4–0.5 °C, which translates into approximately 2% bias in the measured relaxation rates. While this amount of error may be tolerable, it can as well be eliminated by adjusting the setting of variable-temperature unit (in our case, to 36.5–36.6 °C).

Note the results illustrated in Fig. 11 pertain to the sample with moderately high conductivity. In contrast, for the sample in low conductivity buffer (α-spc SH3) the heating generated by spin-lock experiment is only 0.1 °C. For the same sample the PD-CPMG experiment employing 4 kHz DIPSI-2 decoupling and τ_{rel} delays as long as 843 ms produces only 0.2 °C heating.

4. General discussion and concluding remarks

The concepts used in our PD-CPMG experiment are similar to those previously used in the context of relaxation dispersion measurements [50,82]. In particular, Hansen et al. developed a constant-time CW-CPMG experiment where the loss of the signal is measured as a function of n_{tot}^{180}, the total number of 180° ^{15}N pulses. The use of the proton decoupling allows one to focus on the in-phase coherence N_x which has more favorable relaxation properties than the combination of N_x and $2N_yH_z$ previously used in the context of relaxation-compensated CPMG experiment [83]. The

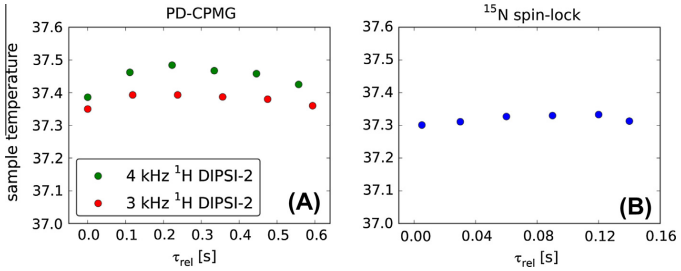


Fig. 11. True temperature associated with different τ_{rel} values as sampled by the PD-CPMG and spin-lock experiments (denatured ubiquitin, 8 M urea, pH 3.0). The same amount of heating is generated when one uses the sample with neutral pH and 50 mM NaCl.

new scheme can also be employed with a smaller number of pulses n_{tot}^{180}, which improves the sampling of relaxation dispersion profiles at low frequency ν_1^{CPMG}. As it should become clear from this brief discussion, the objectives of CW-CPMG experiment are quite different from what is described in this paper, and the technical details are different as well. In fact, it would be difficult to adapt CW-CPMG for the purpose of ^{15}N R_2 measurements since this sequence uses a continuous-wave proton decoupling with a relatively high level of power, 15 kHz.

The new proton-decoupled CPMG experiment has been developed with the goal to improve the precision of ^{15}N R_2 measurements in intrinsically disordered proteins at or near physiological conditions. However, the test measurements described in this paper only partially satisfy these conditions. While ubiquitin data have been collected at 37 °C, the sample was chemically denatured by application of 8 M urea, the pH was 3.0, and no particular attention has been paid to the ionic strength of the solution. How will our experiment fare if the measurements are conducted under more physiologically relevant conditions?

At 37 °C, pH 7.4 amide solvent exchange presents a potentially grave problem. Under these conditions, the exchange rates in disordered proteins mostly fall in the range 20–200 s^{-1} [74], causing significant line broadening and potentially compromising the accuracy of relaxation measurements. Although PD-CPMG measurements under these conditions should be possible, they will be extremely demanding. In this situation it may be advisable to lower pH to 7.0, as this value remains relevant for cytosol [84]. Given that histidine side chains in IDPs titrate near pH 6.0 [85], it should be possible to lower the pH without significantly altering the charges on the protein.

The ion content of the cytosol is typically assumed to be equivalent to 150 mM NaCl. We executed the PD-CPMG experiment on a sample containing this amount of salt and found that the temperature of the sample increased by 1.3 °C. This amount of heating is tolerable and does not affect the appearance of the spectra (verified experimentally). The increase in temperature can be compensated by adjusting the setting of the VT unit. We conclude that PD-CPMG measurements can be conducted using the samples with physiologically relevant salt concentrations.

Finally, let us consider solvent viscosity and the effect of co-solvents. The inside of the cells is known to be crowded, which causes substantial increases in ^{15}N R_2 rates [86]. Similarly, in our ubiquitin measurements the presence of 8 M urea led to elevated R_2 values, both through increased solvent viscosity and coordination of urea to the unfolded protein [87]. In contrast, NMR samples of IDPs are usually prepared without crowding agents or osmolytes. The ^{15}N R_2 rates in such samples are typically found to be 2–3 s^{-1} at the temperatures 15–20 °C [88–90], which corresponds to R_2 of ca. 1 s^{-1} at our target temperature 37 °C. For relaxation decay that is so slow, the advantages of the proposed PD-CPMG scheme should be even more significant. In other words, one may expect gains in precision that would surpass the factor of 3 found in this study.

The issue of precision can be critical for ^{15}N relaxation measurements in IDPs. Because of the propensity to aggregate and precipitate, the samples of IDPs often need to be prepared with low protein concentration, in the range of 10–100 μM. Furthermore, the measurement time is often limited to 1–10 h because the samples tend to quickly deteriorate. Solvent exchange further compounds to the problem, causing line-broadening in the already crowded spectra and further lowering signal-to-noise ratio. Under such circumstances it can be difficult to obtain high-quality data. At the same time it is very important that the data are precise. Indeed, ^{15}N relaxation measurements in IDPs typically pursue the effects that are subtle in nature. For instance, this may be a tenuous α-helical propensity which distinguishes physiologically relevant IDP from a random-coil-like peptide chain or a putative local cluster which is formed in response to a certain point mutation. In each case we rely on small changes in ^{15}N R_2 rates to pinpoint the effect of interest. Thus we are faced with the situation where it is essential, but at the same time very difficult, to collect high-precision ^{15}N R_2 data. The new PD-CPMG experiment is expected to be quite helpful in this regard.

The accuracy of the new experiment has been established through comparison with well-established $R_{1\rho}$ scheme. Such comparison was carried out for both unfolded and folded protein samples. We have determined that PD-CPMG experiment has a number of favorable properties. First, it records pure in-phase evolution during t_1 period and therefore offers an advantage in sensitivity and resolution. The signal-to-noise ratio in the PD-CPMG spectra of α-spc SH3 and ubiquitin are ca. 15% higher than in the comparable $R_{1\rho}$ spectra. This advantage becomes dramatic for proteins or peptides experiencing significant solvent exchange. Second, PD-CPMG data can be easily corrected for ^{15}N offset effect. In fact, this correction is small and can even be neglected, thus obviating the need for companion R_1 experiment. Third, the experiment is designed in such a way that solvent exchange does not interfere with the measurements. We avoid adding D_2O to the sample and maintain the constant degree of water saturation. The contribution from scalar relaxation of the second kind arising from solvent exchange remains minimal. Fourth, the experiment is highly stable with regard to pulse miscalibration and *rf* field inhomogeneity. The amount of sample heating is on par with the $R_{1\rho}$ measurements. Fifth, possible R_{ex} contributions into ^{15}N R_2 rates, which may arise from transient oligomerization or other potentially interesting effects, are largely retained in the PD-CPMG experiment because of the low V_1^{CPMG}. This is arguably preferable to the standard measurement schemes where the application of strong spin-lock field or

frequent ^{15}N 180° pulses leads to partial suppression of R_{ex}. Sixth, PD-CPMG experiment has been adapted to probe $^{15}N^{\varepsilon}$ spins in Arg side chains which are likely to be a valuable target for future studies of protein internal dynamics. We have also verified that PD-CPMG sequence can be employed with ^{13}C-labeled samples using the setting $2\tau_{CP}\sim$5–10 ms (to be reported elsewhere). All of these attractive properties make us believe that PD-CPMG ^{15}N R_2 experiment will become a valuable addition to the arsenal of NMR experiments targeted at intrinsically disordered proteins.

5. Materials and methods

5.1. Protein samples

SH3 domain from chicken α-spectrin and human ubiquitin have been expressed in ^{15}N enriched M9 media and purified as described previously [91,92]. Protein concentrations were determined using UV absorbance at 280 nm. The sample of 1.5 mM α-spc SH3 has been prepared in H_2O, 20 mM citrate, 0.02% NaN_3, pH 3.5. The sample of 2.0 mM unfolded ubiquitin was prepared in H_2O, 8 M urea, pH 3.0. Two additional ubiquitin samples were prepared with protein concentration 50 and 100 μM. Sample volume in each case was 500 μL, with 100 μL of D_2O added into the coaxial insert of the NMR tube (Wilmad, WGS-5BL) to serve as external lock.

5.2. NMR spectroscopy

All NMR experiments were performed on Varian Inova-600 spectrometer equipped with triple-resonance z-axis gradient probe. The measurements were carried out at 25 °C for α-spc SH3 and 37 °C for unfolded ubiquitin. The backbone amide spectra were recorded with ^{15}N carrier at 119 ppm and ^{15}N spectral width of 30 ppm (α-spc SH3) or 22 ppm (ubiquitin). The data were collected with 64 t_1 increments and 1800 points in t_2 domain. The spectra of Arg ($^1H^{\varepsilon}$,$^{15}N^{\varepsilon}$) in α-spc SH3 were recorded with ^{15}N carrier as 85 ppm, ^{15}N spectral width of 10 ppm, and 16 t_1 increments. All data were collected with 4 scans per t_1 increment. The spectra were processed using NMRPipe [93]. Squared phase-shifted sine-bell window function was applied in both dimensions and the spectra were zero-filled to the size 4096×512 ($^1H^N$,^{15}N) or 4096×256 ($^1H^{\varepsilon}$,$^{15}N^{\varepsilon}$).

To ensure a clean comparison between PD-CPMG and spin-lock experiments, we have amended the standard $R_{1\rho}$ sequence [10] by adding a water-crusher element (red box on the right side of Fig. 2). In the spin-lock experiment we have used the following τ_{rel} delays: 5(1), 30(1), 60(2), 90(3), 120(4), and 140(4) ms (shown in brackets is the number of ^1H 180° pulses used to suppress dipolar-CSA cross-correlations). The ^{15}N spin-lock has been applied with *rf* field strength 1.7 kHz. All other experimental settings were the same as in PD-CPMG experiment, including the recycling delay d_1 = 2 s. ^{15}N R_1 data have also been recorded in order to determine ^{15}N offset correction for both $R_{1\rho}$ and PD-CPMG data (cf. Eq. (1)).

The spectra were integrated using the autoFit/nlinLS routines from the NMRPipe package. These routines operate on the entire "stack" of spectral planes, making an assumption that the peak position and the linewidths do not change significantly from one plane to the other and hence can be fitted in a global sense. This approach is especially helpful for quantitation of the weak peaks. The data were fitted assuming Gaussian peak shapes, as appropriate for the squared sine-bell apodization. The spectral assignments of α-spc SH3 and denatured ubiquitin were obtained from the previous reports [94,95].

To monitor sample temperature during the PD-CPMG experiment we have developed a special scheme based on the pulse sequence Fig. 2. Specifically, we changed the original sequence beginning with the acquisition period. Starting from this point the modified sequence reads: $g_{11} - \tau_{blank} - p_{H(90)} - t_2 - d_1$. Here g_{11} is the clean-up gradient with the strength 20 G/cm and duration 3.5 ms, τ_{blank} is 100 ms delay during which time the ^{15}N WALTZ-16 decoupling is turned on, $p_{H(90)}$ is hard proton pulse with the same phase as the receiver, t_2 is the acquisition period of 0.5 s, and d_1 is the recycling delay of 1.5 s. The purpose of this sequence is to record 1D spectrum of methanol (which is contained in the NMR tube coaxial insert) under the conditions that emulate the actual PD-CPMG measurement. We recommend using this accessory experiment when working with high conductivity (high ionic strength) samples, provided that accurate absolute values of R_2 are of interest.

5.3. Numerical simulations

We have simulated spin evolution of the two-spin ($^1H^N$, ^{15}N) and three-spin ($^1H^N$, ^{15}N and $^1H^{\alpha}$) systems during the τ_{rel} period of the PD-CPMG and spin-lock experiments. Taken into consideration were all sources of coherent evolution: *rf* fields, frequency offsets, and scalar couplings. In addition, Redfield matrix has been generated by rigorously including dipolar and CSA contributions. In the 2-spin simulations we have also included the "external" $^1H^{\alpha}$ spin, which manifested itself through dipolar contribution into the auto-relaxation rates of ($^1H^N$, ^{15}N) spin modes. In the 3-spin simulations, $^1H^{\alpha}$ was treated as a part of the spin system. We have assumed that $^3J_{HNH\alpha}$ = 10 Hz, $^2J_{NH\alpha}$ = 5 Hz which is the upper limit for the respective experimentally observed couplings [96,97]. ^{15}N and $^1H^N$ CSA tensors were modeled according to the literature data [8,98]; $^1H^{\alpha}$ CSA tensor was assumed to be zero. The distance between $^1H^N$ and $^1H^{\alpha}$ was set to 1.84 Å. In order to emphasize the effect of dipole–dipole cross-correlations, the angle between ^{15}N–$^1H^N$ and $^1H^N$–$^1H^{\alpha}$ vectors was set to zero. All simulations used a simple form of Lipari-Szabo spectral density [99], where it was assumed that τ_C = 2 ns, S^2 = 0.5, and τ_{fast} = 0. This parameterization is consistent with the previous analyses of ^{15}N relaxation in denatured protein [100] and correctly reproduces typical R_1 and R_2 values observed in denatured ubiquitin at 37 °C (Fig. 3b). *Rf* pulses were simulated with finite length. To model the effect of *rf* field inhomogeneity, the simulations were repeated for 100 discrete values of *rf* field strength uniformly distributed in the interval $[\nu_1 - 2\sigma, \nu_1 + 2\sigma]$. The signal intensity was then calculated as a weighted average of the results from multiple simulations. Solvent exchange was modeled as described previously [49]. Specifically, we recognize that any multi-spin mode combining ^{15}N and $^1H^N$ spin operators undergoes exchange-induced decay with the rate k_{ex}.

The simulations featuring 15×15 and 63×63 propagator matrices were repeated for the same set of τ_{rel} delays as used experimentally. The simulated signal intensities were then fitted with a single exponential, yielding the values of $R_{1\rho}$ or R_2^{app}. The results from 2- and 3-spin simulations proved to be virtually identical. This is not surprising considering that: (i) DIPSI-2 sequence has been designed to perform well in the presence of homonuclear couplings, such as $^3J_{HNH\alpha}$, (ii) the train of ^{15}N 180° CPMG pulses is sufficient to eliminate the effect of the small $^2J_{NH\alpha}$ coupling, with off-resonance DIPSI-2 providing additional insurance, and (iii) three-spin modes generated by dipole–dipole cross-correlations are effectively dephased by DIPSI-2 decoupling.

Acknowledgment

This study was supported by the funds from NSF Grant MCB 1158347.

References

[1] M.E. Hodsdon, D.P. Cistola, Ligand binding alters the backbone mobility of intestinal fatty acid-binding protein as monitored by ^{15}N NMR relaxation and ^{1}H exchange, Biochemistry 36 (1997) 2278–2290.

[2] L. Zidek, M.V. Novotny, M.J. Stone, Increased protein backbone conformational entropy upon hydrophobic ligand binding, Nat. Struct. Biol. 6 (1999) 1118–1121.

[3] M.J. Seewald, K. Pichumani, C. Stowell, B.V. Tibbals, L. Regan, M.J. Stone, The role of backbone conformational heat capacity in protein stability: temperature-dependent dynamics of the B1 domain of Streptococcal protein G, Protein Sci. 9 (2000) 1177–1193.

[4] J.M. Vinther, S.M. Kristensen, J.J. Led, Enhanced stability of a protein with increasing temperature, J. Am. Chem. Soc. 133 (2011) 271–278.

[5] L. Ragona, M. Catalano, M. Luppi, D. Cicero, T. Eliseo, J. Foote, F. Fogolari, L. Zetta, H. Molinari, NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein, J. Biol. Chem. 281 (2006) 9697–9709.

[6] R. Das, M.T. Mazhab-Jafari, S. Chowdhury, S. SilDas, R. Selvaratnam, G. Melacini, Entropy-driven cAMP-dependent allosteric control of inhibitory interactions in exchange proteins directly activated by cAMP, J. Biol. Chem. 283 (2008) 19691–19703.

[7] L.S. Yao, B. Vögeli, J.F. Ying, A. Bax, NMR determination of amide N-H equilibrium bond length from concerted dipolar coupling measurements, J. Am. Chem. Soc. 130 (2008) 16518–16520.

[8] L.S. Yao, A. Grishaev, G. Cornilescu, A. Bax, Site-specific backbone amide ^{15}N chemical shift anisotropy tensors in a small protein from liquid crystal and cross-correlated relaxation measurements, J. Am. Chem. Soc. 132 (2010) 4295–4309.

[9] D.F. Hansen, H.Q. Feng, Z. Zhou, Y.W. Bai, L.E. Kay, Selective characterization of microsecond motions in proteins by NMR relaxation, J. Am. Chem. Soc. 131 (2009) 16257–16265.

[10] D.M. Korzhnev, N.R. Skrynnikov, O. Millet, D.A. Torchia, L.E. Kay, An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates, J. Am. Chem. Soc. 124 (2002) 10743–10753.

[11] F. Ferrage, D. Cowburn, R. Ghose, Accurate sampling of high-frequency motions in proteins by steady-state ^{15}N-{^{1}H} nuclear Overhauser effect measurements in the presence of cross-correlated relaxation, J. Am. Chem. Soc. 131 (2009) 6048–6049.

[12] S. Jurt, O. Zerbe, A study on the influence of fast amide exchange on the accuracy of ^{15}N relaxation rate constants, J. Biomol. NMR 54 (2012) 389–400.

[13] A.M. Mandel, M. Akke, A.G. Palmer, Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme, J. Mol. Biol. 246 (1995) 144–163.

[14] M. Andrec, G.T. Montelione, R.M. Levy, Estimation of dynamic parameters from NMR relaxation data using the Lipari-Szabo model-free approach and Bayesian statistical methods, J. Magn. Reson. 139 (1999) 408–421.

[15] R. Cole, J.P. Loria, FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data, J. Biomol. NMR 26 (2003) 203–213.

[16] D.Q. Jin, M. Andrec, G.T. Montelione, R.M. Levy, Propagation of experimental uncertainties using the Lipari-Szabo model-free analysis of protein dynamics, J. Biomol. NMR 12 (1998) 471–492.

[17] E.J. d'Auvergne, P.R. Gooley, The use of model selection in the model-free analysis of protein dynamics, J. Biomol. NMR 25 (2003) 25–39.

[18] J.H. Chen, C.L. Brooks, P.E. Wright, Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation, J. Biomol. NMR 29 (2004) 243–257.

[19] J.M. Schurr, H.P. Babcock, B.S. Fujimoto, A test of the model-free formulas. Effects of anisotropic rotational diffusion and dimerization, J. Magn. Reson. Ser. B 105 (1994) 211–224.

[20] N. Tjandra, S.E. Feller, R.W. Pastor, A. Bax, Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation, J. Am. Chem. Soc. 117 (1995) 12562–12566.

[21] L.K. Lee, M. Rance, W.J. Chazin, A.G. Palmer, Rotational diffusion anisotropy of proteins from simultaneous analysis of ^{15}N and $^{13}C^{\alpha}$ nuclear spin relaxation, J. Biomol. NMR 9 (1997) 287–298.

[22] P. Dosset, J.C. Hus, M. Blackledge, D. Marion, Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data, J. Biomol. NMR 16 (2000) 23–28.

[23] A.L. Lee, A.J. Wand, Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation, J. Biomol. NMR 13 (1999) 101–112.

[24] S.L. Chang, N. Tjandra, Temperature dependence of protein backbone motion from carbonyl ^{13}C and amide ^{15}N NMR relaxation, J. Magn. Reson. 174 (2005) 43–53.

[25] S.A. Showalter, R. Brüschweiler, Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field, J. Chem. Theory Comput. 3 (2007) 961–975.

[26] S. Mine, S. Tate, T. Ueda, M. Kainosho, T. Imoto, Analysis of the relationship between enzyme activity and its internal motion using nuclear magnetic resonance: ^{15}N relaxation studies of wild-type and mutant lysozyme, J. Mol. Biol. 286 (1999) 1547–1565.

[27] B.F. Volkman, D. Lipson, D.E. Wemmer, D. Kern, Two-state allosteric behavior in a single-domain signaling protein, Science 291 (2001) 2429–2433.

[28] S.W. Homans, Probing the binding entropy of ligand-protein interactions by NMR, ChemBioChem 6 (2005) 1585–1591.

[29] S.R. Tzeng, C.G. Kalodimos, Protein activity regulation by conformational entropy, Nature 488 (2012) 236–240.

[30] A.K. Dunker, C.J. Oldfield, J.W. Meng, P. Romero, J.Y. Yang, J.W. Chen, V. Vacic, Z. Obradovic, V.N. Uversky, The unfoldomics decade: an update on intrinsically disordered proteins, BMC Genomics 9 (S2) (2008) S1.

[31] A.B. Sigalov, A.V. Zhuravleva, V.Y. Orekhov, Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form, Biochimie 89 (2007) 419–421.

[32] J. Song, L.W. Guo, H. Muradov, N.O. Artemyev, A.E. Ruoho, J.L. Markley, Intrinsically disordered gamma-subunit of cGMP phosphodiesterase encodes functionally relevant transient secondary and tertiary structure, Proc. Natl. Acad. Sci. U. S. A 105 (2008) 1505–1510.

[33] M.D. Mukrasch, S. Bibow, J. Korukottu, S. Jeganathan, J. Biernat, C. Griesinger, E. Mandelkow, M. Zweckstetter, Structural polymorphism of 441-residue Tau at single residue resolution, PLoS Biol. 7 (2009) 399–414.

[34] J.A. Jones, P. Hodgkinson, A.L. Barker, P.J. Hore, Optimal sampling strategies for the measurement of spin-spin relaxation times, J. Magn. Reson. Ser. B 113 (1996) 25–34.

[35] P.A. Keifer, 90° degrees pulse width calibrations: how to read a pulse width array, Concept Magn. Reson. 11 (1999) 165–180.

[36] D. Ban, A.D. Gossert, K. Giller, S. Becker, C. Griesinger, D. Lee, Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead, J. Magn. Reson. 221 (2012) 1–4.

[37] F. Massi, E. Johnson, C.Y. Wang, M. Rance, A.G. Palmer, NMR $R_{1\rho}$ rotating-frame relaxation with weak radio frequency fields, J. Am. Chem. Soc. 126 (2004) 2247–2256.

[38] A.G. Palmer, N.J. Skelton, W.J. Chazin, P.E. Wright, M. Rance, Suppression of the effects of cross-correlation between dipolar and anisotropic chemical shift relaxation mechanisms in the measurement of spin spin relaxation rates, Mol. Phys. 75 (1992) 699–711.

[39] S. Zinn-Justin, P. Berthault, M. Guenneugues, H. Desvaux, Off-resonance rf fields in heteronuclear NMR: application to the study of slow motions, J. Biomol. NMR 10 (1997) 363–372.

[40] Y. Pang, M. Buck, E.R.P. Zuiderweg, Backbone dynamics of the ribonuclease binase active site area using multinuclear (^{15}N and ^{13}CO) NMR relaxation and computational molecular dynamics, Biochemistry 41 (2002) 2655–2666.

[41] W. Myint, Q.G. Gong, R. Ishima, Practical aspects of ^{15}N CPMG transverse relaxation experiments for proteins in solution, Concept Magn. Reson. A 34A (2009) 63–75.

[42] G.N.B. Yip, E.R.P. Zuiderweg, A phase cycle scheme that significantly suppresses offset-dependent artifacts in the R_2-CPMG ^{15}N relaxation experiment, J. Magn. Reson. 171 (2004) 25–36.

[43] D. Long, M.L. Liu, D.W. Yang, Accurately probing slow motions on millisecond timescales with a robust NMR relaxation experiment, J. Am. Chem. Soc. 130 (2008) 2432–2433.

[44] A.D. Bain, C.K. Anand, Z.H. Nie, Exact solution of the CPMG pulse sequence with phase variation down the echo train: application to R_2 measurements, J. Magn. Reson. 209 (2011) 183–194.

[45] W. Myint, Y.F. Cai, C.A. Schiffer, R. Ishima, Quantitative comparison of errors in ^{15}N transverse relaxation rates measured using various CPMG phasing schemes, J. Biomol. NMR 53 (2012) 13–23.

[46] A.J. Shaka, C.J. Lee, A. Pines, Iterative schemes for bilinear operators: application to spin decoupling, J. Magn. Reson. 77 (1988) 274–293.

[47] A. Bax, M. Ikura, L.E. Kay, D.A. Torchia, R. Tschudin, Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins, J. Magn. Reson. 86 (1990) 304–318.

[48] T.J. Norwood, J. Boyd, J.E. Heritage, N. Soffe, I.D. Campbell, Comparison of techniques for ^{1}H-detected heteronuclear ^{1}H–^{15}N spectroscopy, J. Magn. Reson. 87 (1990) 488–501.

[49] N.R. Skrynnikov, R.R. Ernst, Detection of intermolecular chemical exchange through decorrelation of two-spin order, J. Magn. Reson. 137 (1999) 276–280.

[50] D.F. Hansen, P. Vallurupalli, L.E. Kay, An improved ^{15}N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins, J. Phys. Chem. B 112 (2008) 5898–5904.

[51] B.A. Messerle, G. Wider, G. Otting, C. Weber, K. Wüthrich, Solvent suppression using a spin lock in 2D and 3D NMR spectroscopy with H_2O solutions, J. Magn. Reson. 85 (1989) 608–613.

[52] L.E. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, J. Am. Chem. Soc. 114 (1992) 10663–10665.

[53] J. Schleucher, M. Sattler, C. Griesinger, Coherence selection by gradients without signal attenuation: application to the 3-dimensional HNCO experiment, Angew. Chem.-Int. Ed. Engl. 32 (1993) 1489–1491.

[54] D. Marion, M. Ikura, R. Tschudin, A. Bax, Rapid recording of 2D NMR spectra without phase cycling: application to the study of hydrogen exchange in proteins, J. Magn. Reson. 85 (1989) 393–399.

[55] E. Kupce, J. Boyd, I.D. Campbell, Short selective pulses for biochemical applications, J. Magn. Reson. Ser. B 106 (1995) 300–303.

[56] H. Geen, R. Freeman, Band-selective radiofrequency pulses, J. Magn. Reson. 93 (1991) 93–141.

[57] S. Grzesiek, A. Bax, The importance of not saturating H_2O in Protein NMR: application to sensitivity enhancement and NOE measurements, J. Am. Chem. Soc. 115 (1993) 12593–12594.

[58] N.A. Farrow, R. Muhandiram, A.U. Singer, S.M. Pascal, C.M. Kay, G. Gish, S.E. Shoelson, T. Pawson, J.D. Forman-Kay, L.E. Kay, Backbone dynamics of a free and a phosphopeptide-complexed src homology 2 domain studied by ^{15}N NMR relaxation, Biochemistry 33 (1994) 5984–6003.

[59] K. Chen, N. Tjandra, Water proton spin saturation affects measured protein backbone ^{15}N spin relaxation rates, J. Magn. Reson. 213 (2011) 151–157.

[60] N.A. Lakomek, J.F. Ying, A. Bax, Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods, J. Biomol. NMR 53 (2012) 209–221.

[61] A.E. Kelly, H.D. Ou, R. Withers, V. Dotsch, Low-conductivity buffers for high-sensitivity NMR measurements, J. Am. Chem. Soc. 124 (2002) 12013–12019.

[62] F. Lohr, H. Ruterjans, Detection of nitrogen-nitrogen J-couplings in proteins, J. Magn. Reson. 132 (1998) 130–137.

[63] R. Ishima, D.A. Torchia, Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach, J. Biomol. NMR 25 (2003) 243–248.

[64] A.L. Hansen, P. Lundstrom, A. Velyvis, L.E. Kay, Quantifying millisecond exchange dynamics in proteins by CPMG relaxation dispersion NMR using side-chain ^{1}H probes, J. Am. Chem. Soc. 134 (2012) 3178–3189.

[65] D.M. Korzhnev, E.V. Tischenko, A.S. Arseniev, Off-resonance effects in ^{15}N T_2 CPMG measurements, J. Biomol. NMR 17 (2000) 231–237.

[66] G.S. Rule, T.K. Hitchens, Fundamentals of Protein NMR Spectroscopy, Springer, Netherlands, 2006.

[67] M. Kjaergaard, S. Brander, F.M. Poulsen, Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH, J. Biomol. NMR 49 (2011) 139–149.

[68] J. Iwahara, Y.S. Jung, G.M. Clore, Heteronuclear NMR spectroscopy for lysine NH3 groups in proteins: unique effect of water exchange on ^{15}N transverse relaxation, J. Am. Chem. Soc. 129 (2007) 2971–2980.

[69] B.S. Blaum, J.A. Deakin, C.M. Johansson, A.P. Herbert, P.N. Barlow, M. Lyon, D. Uhrin, Lysine and arginine side chains in glycosaminoglycan-protein complexes investigated by NMR, cross-linking, and mass spectrometry: a case study of the Factor H - Heparin interaction, J. Am. Chem. Soc. 132 (2010) 6374–6381.

[70] K. Hayamizu, W.S. Price, A new type of sample tube for reducing convection effects in PGSE-NMR measurements of self-diffusion coefficients of liquid samples, J. Magn. Reson. 167 (2004) 328–333.

[71] T. Yamazaki, S.M. Pascal, A.U. Singer, J.D. Forman-Kay, L.E. Kay, NMR pulse schemes for the sequence-specific assignment of arginine guanidino ^{15}N and ^{1}H chemical shifts in proteins, J. Am. Chem. Soc. 117 (1995) 3556–3564.

[72] J.C. Hindman, A. Svirmickas, M. Wood, Relaxation processes in water. A study of the proton spin-lattice relaxation time, J. Chem. Phys. 59 (1973) 1517–1522.

[73] A. Abragam, Principles of Nuclear Magnetism, Oxford University Press, USA, 1983.

[74] Y.W. Bai, J.S. Milne, L. Mayne, S.W. Englander, Primary structure effects on peptide group hydrogen exchange, Proteins: Struct., Funct., Genet. 17 (1993) 75–86.

[75] F. Kateb, P. Pelupessy, G. Bodenhausen, Measuring fast hydrogen exchange rates by NMR spectroscopy, J. Magn. Reson. 184 (2007) 108–113.

[76] P.R. Vasos, J.B. Hall, R. Kummerle, D. Fushman, Measurement of ^{15}N relaxation in deuterated amide groups in proteins using direct nitrogen detection, J. Biomol. NMR 36 (2006) 27–36.

[77] A.C. Wang, A. Bax, Minimizing the effects of radiofrequency heating in multidimensional NMR experiments, J. Biomol. NMR 3 (1993) 715–720.

[78] V.Y. Orekhov, D.M. Korzhnev, T. Diercks, H. Kessler, A.S. Arseniev, ^{1}H–^{15}N NMR dynamic study of an isolated alpha-helical peptide (1–36)-bacteriorhodopsin reveals the equilibrium helix-coil transitions, J. Biomol. NMR 14 (1999) 345–356.

[79] C. Ammann, P. Meier, A.E. Merbach, A simple multi-nuclear NMR thermometer, J. Magn. Reson. 46 (1982) 319–321.

[80] J. Wirmer, W. Peti, H. Schwalbe, Motional properties of unfolded ubiquitin: a model for a random coil protein, J. Biomol. NMR 35 (2006) 175–186.

[81] Y. Xue, N.R. Skrynnikov, Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, ^{15}N relaxation, and Molecular Dynamics simulations: how fast is segmental diffusion in denatured ubiquitin?, J Am. Chem. Soc. 133 (2011) 14614–14628.

[82] S.M. Kristensen, G. Siegal, A. Sankar, P.C. Driscoll, Backbone dynamics of the C-terminal SH2 domain of the p85 alpha subunit of phosphoinositide 3-kinase: effect of phosphotyrosine-peptide binding and characterization of slow conformational exchange processes, J. Mol. Biol. 299 (2000) 771–788.

[83] J.P. Loria, M. Rance, A.G. Palmer, A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy, J. Am. Chem. Soc. 121 (1999) 2331–2332.

[84] R. Orij, J. Postmus, A. Ter Beek, S. Brul, G.J. Smits, In vivo measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in Saccharomyces cerevisiae reveals a relation between intracellular pH and growth, Microbiology 155 (2009) 268–278.

[85] R.L. Croke, S.M. Patil, J. Quevreaux, D.A. Kendall, A.T. Alexandrescu, NMR determination of pK_a values in α-synuclein, Protein Sci. 20 (2011) 256–269.

[86] E.A. Cino, M. Karttunen, W.Y. Choy, Effects of molecular crowding on the dynamics of intrinsically disordered proteins, PLoS One 7 (2012) e49876.

[87] M. Candotti, S. Esteban-Martin, X. Salvatella, M. Orozco, Toward an atomistic description of the urea-denatured state of proteins, Proc. Natl. Acad. Sci. U. S. A 110 (2013) 5933–5938.

[88] M.K. Cho, G. Nodet, H.Y. Kim, M.R. Jensen, P. Bernado, C.O. Fernandez, S. Becker, M. Blackledge, M. Zweckstetter, Structural characterization of alpha-synuclein in an aggregation prone state, Protein Sci. 18 (2009) 1840–1846.

[89] R. Silvers, F. Sziegat, H. Tachibana, S. Segawa, S. Whittaker, U.L. Gunther, F. Gabel, J.R. Huang, M. Blackledge, J. Wirmer-Bartoschek, H. Schwalbe, Modulation of structure and dynamics by disulfide bond formation in unfolded states, J. Am. Chem. Soc. 134 (2012) 6846–6854.

[90] J. Roche, J. Ying, A.S. Maltsev, A. Bax, Impact of hydrostatic pressure on an intrinsically disordered protein: a high-pressure NMR study of α-synuclein, Chembiochem 14 (2013) 1754–1761.

[91] J. Pauli, B. van Rossum, H. Forster, H.J.M. de Groot, H. Oschkinat, Sample optimization and identification of signal patterns of amino acid side chains in 2D RFDR spectra of the α-spectrin SH3 domain, J. Magn. Reson. 143 (2000) 411–416.

[92] G.A. Lazar, J.R. Desjarlais, T.M. Handel, De novo design of the hydrophobic core of ubiquitin, Protein Sci. 6 (1997) 1167–1178.

[93] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe – a multidimensional spectral processing system based on unix pipes, J. Biomol. NMR 6 (1995) 277–293.

[94] F.J. Blanco, A.R. Ortiz, L. Serrano, ^{1}H and ^{15}N NMR assignment and solution structure of the SH3 domain of spectrin: comparison of unrefined and refined structure sets with the crystal structure, J. Biomol. NMR 9 (1997) 347–357.

[95] W. Peti, L.J. Smith, C. Redfield, H. Schwalbe, Chemical shifts in denatured proteins: Resonance assignments for denatured ubiquitin and comparisons with other denatured proteins, J. Biomol. NMR 19 (2001) 153–165.

[96] A.C. Wang, A. Bax, Reparameterization of the Karplus relation for ^{3}J(H^{α}-N) and ^{3}J(H^{N}-C') in peptides from uniformly ^{13}C/^{15}N-enriched human ubiquitin, J. Am. Chem. Soc. 117 (1995) 1810–1813.

[97] J.M. Schmidt, Y. Hua, F. Lohr, Correlation of ^{2}J couplings with protein secondary structure, Proteins 78 (2010) 1544–1562.

[98] M. Ottiger, N. Tjandra, A. Bax, Magnetic field dependent amide ^{15}N chemical shifts in a protein-DNA complex resulting from magnetic ordering in solution, J. Am. Chem. Soc. 119 (1997) 9825–9830.

[99] G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, J. Am. Chem. Soc. 104 (1982) 4546–4559.

[100] N.A. Farrow, O.W. Zhang, J.D. Forman-Kay, L.E. Kay, Characterization of the backbone dynamics of folded and denatured states of an SH3 domain, Biochemistry 36 (1997) 2390–2402.

J Biomol NMR (2014) 58:175–192
DOI 10.1007/s10858-014-9815-5

ARTICLE

CP-HISQC: a better version of HSQC experiment for intrinsically disordered proteins under physiological conditions

Tairan Yuwen · Nikolai R. Skrynnikov

Received: 18 November 2013 / Accepted: 24 January 2014 / Published online: 5 February 2014

Abstract ^{1}H–^{15}N HSQC spectroscopy is a workhorse of protein NMR. However, under physiological conditions the quality of HSQC spectra tends to deteriorate due to fast solvent exchange. For globular proteins only a limited number of surface residues are affected, but in the case of intrinsically disordered proteins (IDPs) HSQC spectra are thoroughly degraded, suffering from both peak broadening and loss of intensity. To alleviate this problem, we make use of the following two concepts. (1) Proton-decoupled HSQC. Regular HSQC and its many variants record the evolution of multi-spin modes, $2N_xH_z$ or $2N_xH_x$, in indirect dimension. Under the effect of fast solvent exchange these modes undergo rapid decay, which results in severe line-broadening. In contrast, proton-decoupled HSQC relies on N_x coherence which is essentially insensitive to the effects of solvent exchange. Moreover, for measurements involving IDPs at or near physiological temperature, N_x mode offers excellent relaxation properties, leading to very sharp resonances. (2) Cross-polarization ^{1}H-to-^{15}N transfer. If CP element is designed such as to lock both $^{1}H^{N}$ and water magnetization, the following transfer is effected: $H_x^{water} \rightarrow H_x^N \rightarrow N_x$. Thus water magnetization is successfully exploited to boost the amount of signal. In addition, CP element suffers less loss from solvent exchange, conformational exchange, and dipolar relaxation compared to the more popular INEPT element. Combining these two concepts, we have implemented the experiment termed CP-HISQC (*c*ross-*p*olarization assisted *h*eteronuclear *i*n-phase *s*ingle-*q*uantum *c*orrelation). The pulse sequence has been designed such as to preserve water magnetization and therefore can be executed with reasonably short recycling delays. In the presence of fast solvent exchange, $k_{ex} \sim 100\ s^{-1}$, CP-HISQC offers much better spectral resolution than conventional HSQC-type experiments. At the same time it offers up to twofold gain in sensitivity compared to plain proton-decoupled HSQC. The new sequence has been tested on the sample of drkN SH3 domain at pH 7.5, 30 °C. High-quality spectrum has been recorded in less than 1 h, containing resonances from both folded and unfolded species. High-quality spectra have also been obtained for arginine side-chain $H^{\varepsilon}N^{\varepsilon}$ groups in the sample of short peptide Sos. For Arg side chains, we have additionally implemented (HE)NE(CD)HD experiment. Using ^{13}C-labeled sample of Sos, we have demonstrated that proton-to-nitrogen CP transfer remains highly efficient in the presence of solvent exchange as fast as $k_{ex} = 620\ s^{-1}$. In contrast, INEPT transfer completely fails in this regime.

Keywords ^{1}H–^{15}N correlation spectroscopy · Proton-decoupled HSQC · Cross-polarization · Amide solvent exchange · Intrinsically disordered proteins · Arginine side-chain $H^{\varepsilon}N^{\varepsilon}$ group

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9815-5) contains supplementary material, which is available to authorized users.

T. Yuwen · N. R. Skrynnikov (✉)
Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA
e-mail: nikolai@purdue.edu

N. R. Skrynnikov
Laboratory of Biomolecular NMR, St. Petersburg State University, St. Petersburg 199034, Russia

Introduction

Two-dimensional ^{1}H–^{15}N correlation spectroscopy is the most widely used class of experiments in the field of protein NMR. There are a number of reasons for continued popularity of these experiments. Bacterially expressed recombinant proteins with ^{15}N labeling can be produced using conventional protocols at very little extra cost. HSQC-type experiments generally have high sensitivity and yield the spectra that are well dispersed and free from interference of the water signal. Furthermore, these spectra are exquisitely sensitive to changes in protein status, such as ligand binding or conformational rearrangements. The HSQC-based sequences are also used to measure ^{15}N relaxation rates, paramagnetic relaxation enhancements, residual dipolar couplings, and other experimental parameters of prime importance.

For all of its strengths, ^{1}H–^{15}N correlation spectroscopy is not ideally suited for measurements at or near physiological conditions (pH 7.4, 37 °C). Under these conditions the quality of the spectra deteriorates due to rapid $^{1}H^{N}$ solvent exchange which causes both loss of intensity and line broadening. In the case of folded soluble proteins only those amides that lie on the surface of the protein and have a significant degree of solvent exposure are affected. However, among these affected residues one often finds key catalytic residues, as well as residues which are critically important for ligand binding (Mueller et al. 2001; Amezcua et al. 2002; Haddad et al. 2005). A similar situation is encountered in the samples of membrane proteins reconstituted in lipids or detergents; in this case solvent exchange has an effect on water-exposed loops (Fernandez et al. 2004). Of particular importance for our study, solvent exchange can have severe adverse impact on ^{1}H–^{15}N spectroscopy of intrinsically disordered proteins (IDPs). The degree of water exposure is typically very high for most residues in IDPs. As a consequence, already at room temperature and pH 6.0 the quality of the spectra visibly deteriorates. This happens regardless of residual secondary structure. Indeed, the presence of transient, low-populated α-helices or β-turns offers little protection against solvent exchange (Croke et al. 2008; Hsu et al. 2009; Gray et al. 2012). Even in the case of molten globule proteins, solvent exchange can thoroughly degrade the quality of ^{1}H–^{15}N spectra (Loh et al. 1995).

In principle, the problems associated with solvent exchange can be avoided if one resorts to pulse sequences employing non-labile protons or heteronuclei. A number of experiments have been developed along these lines including, notably, ^{13}C-based experiments (Bermel et al. 2006; Pasat et al. 2008; Felli and Brutscher 2009; Werbeck et al. 2013). However, given the substantial cost of ^{13}C-labeling and low sensitivity of the proton-less pulse schemes, the ^{1}H–^{15}N spectroscopy remains the first choice for NMR studies of IDPs.

It has been recognized early on that in the presence of moderately fast solvent exchange it is important to preserve water magnetization (Grzesiek and Bax 1993; Farrow et al. 1994). Along these lines van Zijl and co-workers developed FHSQC sequence, which offers excellent water preservation properties (Mori et al. 1995). Transfer of magnetization from water to labile protein sites makes it possible to relatively quickly replenish $^{1}H^{N}$ magnetization. As a result, FHSQC experiment can be executed with relatively short recycling delays. The same is true for the more recent SOFAST-HMQC experiment (Schanda and Brutscher 2005). The superior performance of SOFAST-HMQC under moderately fast exchange conditions, $k_{ex} \sim 1$–$10\ s^{-1}$, has been recently documented (Yao et al. 2011).

While the experiments such as FHSQC and SOFAST-HMQC offer superior sensitivity, they also suffer from one potential setback. Specifically, these sequences utilize multispin modes such as $2N_xH_z$ or $2N_xH_x$ to record the evolution in t_1 domain. Under the effect of solvent exchange, such multispin modes undergo rapid decay, resulting in dramatic line broadening (Skrynnikov and Ernst 1999). As a result, the increased solvent exchange leads to loss of spectral resolution in nitrogen dimension.

This problem has been addressed in another pulse sequence, which has been originally termed proton-decoupled HSQC (Bax et al. 1990). This experiment relies on N_x coherence to record the evolution in t_1 domain. In the presence of robust proton decoupling, N_x evolves essentially independently of solvent exchange. In the case of highly mobile disordered protons, the spectra can be recorded with very narrow linewidth in ^{15}N dimension. The favorable properties of proton-decoupled HSQC with respect to solvent exchange have been pointed out already in the original report (Bax et al. 1990). Later the same strategy has been used to record the spectra of lysine side-chain NH_3 groups, which are otherwise difficult to observe due to the fast solvent exchange (Iwahara et al. 2007). The lysine experiment was termed *h*eteronuclear *i*n-phase *s*ingle *q*uantum *c*oherence spectroscopy or HISQC. Very recently, we have used this scheme in the context of ^{15}N R_2 measurements in IDPs (Yuwen and Skrynnikov 2013).

In this report we present the proton-decoupled HSQC experiment which uses cross-polarization (CP) element as a starting point. The new pulse sequence is termed CP-HISQC. As demonstrated below, the CP element makes efficient use of water magnetization, achieving the following transfer:

$$H_x^{water} \xrightarrow{k_{ex}} H_x^N \xrightarrow{^1J_{HN}} N_x \quad (1)$$

For the sample conditions close to physiological the use of the CP scheme affords significant gain in sensitivity

compared to the previously reported HISQC sequence. At the same time, CP-HISQC offers dramatically improved spectral resolution compared to conventional experiments such as FHSQC and SOFAST-HMQC. The combination of superior resolution and good sensitivity makes CP-HISQC sequence particularly well-suited for studies of IDPs at or near physiological conditions. It can also be useful for studies of globular proteins (membrane proteins) if the goal is to capture the signals from surface sites undergoing rapid solvent exchange.

Setting up the experiment is sufficiently easy, and the efficient CP transfer is achieved simultaneously for those residues that experience fast solvent exchange and for those residues that are protected. This property has been demonstrated on the sample of N-terminal SH3 domain of adaptor protein drk, which contains both folded and unfolded forms of the protein (Farrow et al. 1995). Using CP-HISQC sequence we have recorded a high-quality backbone $^{1}H^{N}$–^{15}N spectrum of drkN SH3 at 30 °C, pH 7.5. In addition, we have also recorded arginine side-chain $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ spectra of ten-residue peptide from guanine nucleotide exchange factor Sos (Wu et al. 1995). These data were collected at pH 6.0 in the range of temperatures from 23 to 37 °C. Using the sample of Sos peptide we have also shown that CP transfer remains efficient even in the presence of very fast solvent exchange (50 °C, $k_{ex} \sim 600\ s^{-1}$).

Results

CP-HISQC pulse sequence

The new pulse sequence CP-HISQC (*c*ross-*p*olarization assisted *h*eteronuclear *i*n-phase *s*ingle-*q*uantum *c*orrelation) is shown in Fig. 1. The initial hard ^{1}H 90° pulse followed by high-power DIPSI-2 pulse train lock both amide and water magnetization along the x axis. As amides undergo solvent exchange, the H_x magnetization on amides is gradually replaced with H_x magnetization from water. At the same time H_x is continuously transformed into N_x via the cross-polarization transfer. The ability of the CP

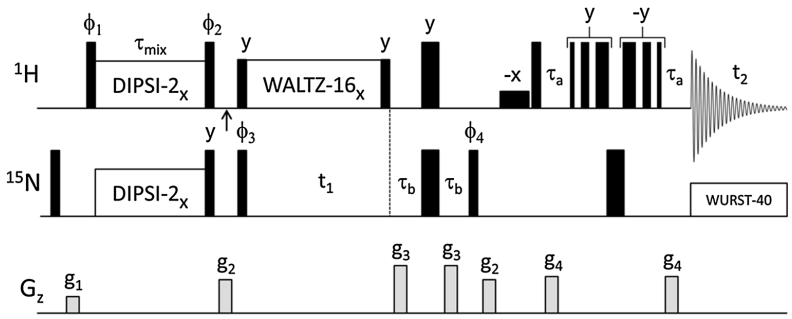


Fig. 1 CP-HISQC pulse sequence for ^{1}H–^{15}N correlation spectroscopy. The parameters below are for backbone amide measurements using the sample of drkN SH3 (unless indicated otherwise). Proton carrier is initially placed at the middle of the amide region, 8.20 ppm. At the point indicated by vertical arrow, the carrier is jumped to water resonance, 4.71 ppm. Hard 90° proton pulses have duration 6.1 μs. 90° water flip-back pulse has rectangular shape and duration 880 μs. DIPSI-2 cross-polarization element (Rucker and Shaka 1989) uses 4.0 kHz *rf* field on both channels (corresponding to 90° pulses of duration 62 μs). WALTZ-16 decoupling (Shaka et al. 1983) is applied with *rf* field strength 5.0 kHz (the same as for the flanking 90° pulses). This is sufficient to minimize the loss of N_x coherence due to solvent exchange (scalar relaxation) (Yuwen and Skrynnikov 2013). Nitrogen carrier has been set at 119 ppm. Hard ^{15}N pulses are of duration 38 μs. WURST-40 decoupling (Kupce and Freeman 1996) during the acquisition period uses *rf* field strength 0.6 kHz. The durations of delays $\tau_a = 0.8 \times 1/(4J_{NH}) = 2.15$ ms and $\tau_b = 0.9 \times 1/(4J_{NH}) = 2.40$ ms have been optimized based on the model calculations assuming $k_{ex} = 100\ s^{-1}$. The duration of the CP element τ_{mix} is $(3/2) \times T_{DIPSI2}$, where T_{DIPSI2} is the length of full DIPSI-2 cycle corresponding to 10360° rotation. This translates into $\tau_{mix} = 10.705$ ms, as used in the actual experiment. The spacing between the consecutive pulses in 3-9-19 WATERGATE (Sklenar et al. 1993) is 183 μs. The phase cycle employed is: ϕ_1 = y, −y; ϕ_2 = −y, y; ϕ_3 = 2(y), 2(−y); ϕ_4 = 4(x), 4(−x); ϕ_{rec} = x, −x, −x, x, −x, x, x, −x. DIPSI-2 and WALTZ-16 sequences are applied with (initial) phase x. The TPPI scheme is implemented by inverting ϕ_4 in concert with ϕ_{rec} (Marion et al. 1989). Quadrature detection in F_1 is achieved by incrementing ϕ_4 by π/2. Four scans per t_1 increment are sufficient to obtain high-quality spectra free from any artefacts. Gradient strengths in G/cm (length in ms) are: $g_1 = 5.0$ (1.0), $g_2 = 10.0$ (1.0), $g_3 = 15.0$ (1.25), $g_4 = 11.0$ (1.0). We choose not to employ the enhanced sensitivity scheme (Palmer et al. 1991; Kay et al. 1992) since it requires ca. extra 5 ms transfer time, which translates into significant loss of signal in the presence of solvent exchange with $k_{ex} \sim 100\ s^{-1}$. When recording $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ correlation map for Arg side chains, proton *rf* carrier is initially placed in $^{1}H^{\varepsilon}$ region, 7.2 ppm, and then jumped to water. Nitrogen carrier is positioned at 85 ppm. The $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ spectra have been recorded using ^{13}C-labeled sample of peptide Sos. To suppress J_{NC} evolution, a composite $(90^{\circ}_{y}180^{\circ}_{x}90^{\circ}_{y})$ carbon pulse has been inserted in the middle of the t_1 evolution period (net duration 66 μs, *rf* carrier at 118 ppm). Other settings, including the CP element, are the same as in the backbone experiment

element to efficiently utilize the influx of new water magnetization is the key to its superior performance.

Following the CP element, N_x magnetization is ^{15}N-frequency-labeled during the evolution period t_1, while proton decoupling is applied to suppress $^1J_{NH}$ and thus prevent the build-up of anti-phase magnetization. The use of the in-phase nitrogen magnetization greatly improves the quality of the spectra compared to the standard scheme (where $2N_xH_z$ evolves during t_1, suffering dramatic losses due to rapid solvent exchange). Following the evolution period, refocused INEPT is used to return magnetization to amide protons for detection. While CP scheme is strongly preferred for the initial proton-to-nitrogen transfer, the reverse transfer is best accomplished using INEPT (the practical considerations dictating this choice are discussed below).

Special care has been taken to preserve water magnetization throughout the pulse sequence. As already pointed out, water magnetization is locked during the CP element; likewise, it is locked during t_1 period by means of the WALTZ sequence of phase x. Finally, during the refocused INEPT water magnetization is maintained along the z axis. As a result, at the end of each scan water retains ca. 75 % of its equilibrium magnetization, which is then quickly transferred to amide sites through solvent exchange. In this manner we avoid long recycling delays which would otherwise be necessary to allow for recovery of water magnetization.

The sequence shown in Fig. 1 is similar to the well-known proton-decoupled HSQC experiment (Bax et al. 1990), as well as more recent HISQC experiment (Iwahara et al. 2007) developed for Lys side-chain NH_3 groups. The main distinction of our sequence is the use of the CP element which leads to significantly better sensitivity (see below). We have also found that 3-9-19 WATERGATE (Sklenar et al. 1993) offers better water preservation compared to the version employing 90° selective rectangular pulses (Iwahara et al. 2007).

The CP element has been designed such as to lock both water and $^1H^N$ magnetization along the x axis. We have determined that optimal results are obtained when (i) carrier is set on $^1H^N$ and (ii) the strength of DIPSI *rf* field is at least two times greater than the frequency offset between $^1H^N$ and water, $\nu_1^H/\Delta \geq 2.0$. In principle, increasing the strength of *rf* field beyond this level is useful, but the gains tend to level off (discussed in what follows).

Strictly speaking, the *rf* field strength $\nu_1^H = 2\Delta$ is insufficient to properly align water magnetization along the x axis. Following the initial hard 90° pulse, water magnetization lands on the x axis. Subsequently, under the effect of off-resonance DIPSI, water magnetization travels on a complicated trajectory on the surface of a unit sphere (while maintaining a significant x projection). This situation can be described in terms of water magnetization being partially locked. After full DIPSI-2 cycle is completed, water is returned to the x axis, reflecting excellent offset compensation properties of the DIPSI-2 composite pulse. At this point the second hard 90° pulse is applied, returning water magnetization to the z axis. This scheme ensures good preservation of water magnetization, which is essential for the sensitivity of the measurements.

The necessary condition for this favorable scenario is that the length of the CP period, τ_{mix}, is proportional to the integer number of DIPSI-2 cycles. In practice, we have found that good results are obtained already with semi-integer number of DIPSI-2 cycles, $\tau_{mix} = (n/2) \times T_{DIPSI2}$. On the other hand, the length of τ_{mix} is dictated by the requirements of cross-polarization transfer. In the absence of solvent exchange, the optimal transfer is achieved when $\tau_{mix} = 1/J_{NH} = 10.8$ ms. How to reconcile these two requirements as per the length of τ_{mix}?

It turns out that in the presence of rapid solvent exchange high efficiency of the CP transfer is attained in a broad interval of τ_{mix}. Specifically, the range of τ_{mix} ca. 10–13 ms can be recommended (see Fig. 5 below). Within this range it is easy to accommodate the semi-integer number of DIPSI-2 cycles. For example, in our backbone drkN SH3 experiment we have used $\tau_{mix} = (3/2) \times T_{DIPSI2} = 10.705$ ms.

CP-HISQC spectra of $^{15}N^{\varepsilon}$–$^1H^{\varepsilon}$ groups from Arg side chains

Sos peptide with the primary sequence PPPVPPRRRR is a proline-rich fragment from guanine nucleotide exchange factor Sos. This fragment is responsible for the interaction with SH3 domains of various adaptor proteins (Wu et al. 1995). The signals from four arginine $^{15}N^{\varepsilon}$–$^1H^{\varepsilon}$ groups in Sos are largely overlapped; the spectrum contains two resolved peaks of unequal height (see Fig. 2, bottom row).

Solvent exchange rates for $^{15}N^{\varepsilon}$–$^1H^{\varepsilon}$ sites can be measured by means of the experiment that records $^{13}C^{\delta}$,$^1H^{\delta}$ correlation map (Segawa et al. 2008). In the case of Sos peptide, this spectrum features a single four-fold degenerate peak. As a consequence, k_{ex} constant obtained in this experiment represents an effective average of the four exchange rates. Given that four arginines are contiguously located at the C-terminus of the Sos peptide and are dynamically disordered, it is reasonable to assume that their individual exchange rates are all similar to each other and close to the experimentally measured k_{ex}. Using ^{13}C-labeled sample of Sos with pH 6.0 and pulse sequence adapted from (Segawa et al. 2008) we determined $^{15}N^{\varepsilon}$–$^1H^{\varepsilon}$ solvent exchange rate at three different temperatures: 43 s^{-1} at 23 °C, 86 s^{-1} at 30 °C, and 170 s^{-1} at 37 °C.

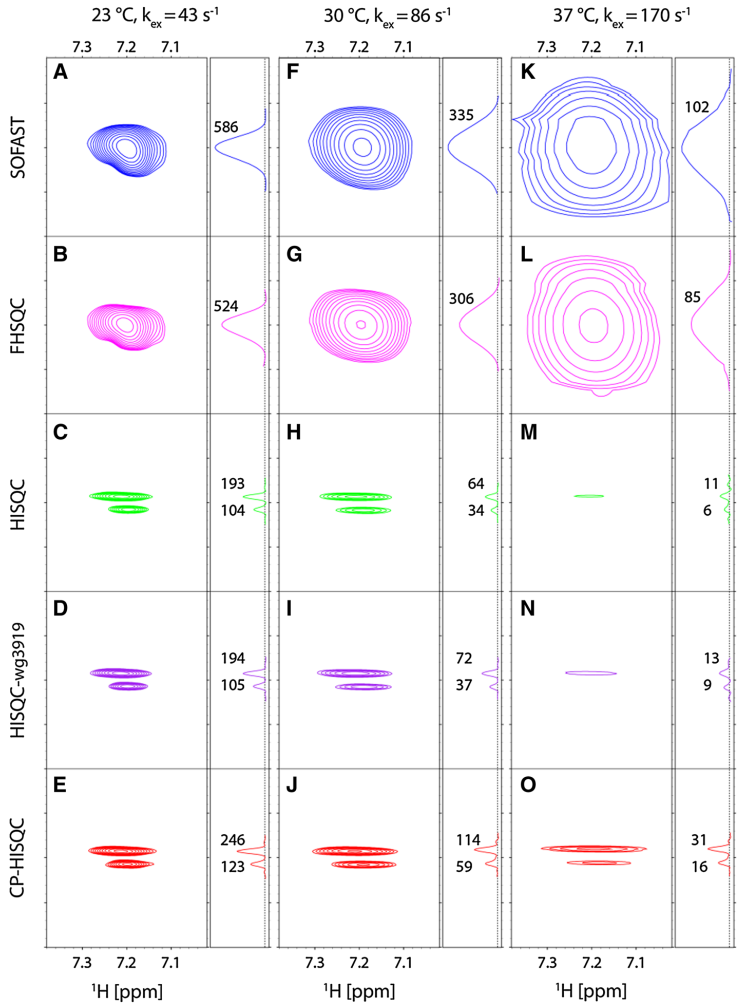


Fig. 2 Arginine $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ correlation maps obtained for the sample of Sos peptide at three different temperatures using five different experimental schemes (as indicated in the plot). The durations of all experiments are approximately equal, ca. 10 min. The *side panels* show 1D cross-sections at 7.2 ppm; the number next to each peak indicates the corresponding S/N ratio. All spectra taken at the same temperature are plotted with the same scale. The separation between the two $^{15}N^{\varepsilon}$ resonances is 0.3 ppm

This range of k_{ex} is well suited for testing the performance of the new CP-HISQC experiment.

The $^{15}N^{\varepsilon}$–$^{1}H^{\varepsilon}$ spectral maps from the new CP-HISQC experiment have been compared with the spectra from four other popular experiments with optimized sensitivity and/or resolution: SOFAST-HMQC (Schanda et al. 2005), FHSQC (Mori et al. 1995), HISQC (Iwahara et al. 2007), and HISQC-wg3919 (our modification of the original HISQC sequence utilizing 3-9-19 WATERGATE element). The latter three experiments are specifically designed for samples with fast solvent exchange. All spectra were recorded using the recycling delays optimized for best signal-to-noise ratio: 0.2 s for SOFAST-HMQC and FHSQC, 1.0 s for HISQC, HISQC-wg3919, and CP-HISQC. When choosing the recycling delay for CP-HISQC, one should bear in mind that cross-polarization element creates a certain load on the probe (4.0 kHz ^{15}N *rf* field applied for 10.7 ms), although this is not a factor for the latest-generation probes which can easily handle such amount of power while requiring only short recycling delays (Ban et al. 2012). All experiments illustrated in Fig. 2 use the same INEPT delays, which have been optimized based on model calculations assuming $k_{ex} = 100\ s^{-1}$.

The data in Fig. 2 show that the SOFAST-HMQC and FHSQC sequences which rely on the evolution of two-spin coherences ($2N_xH_x$ and $2N_xH_z$, respectively) suffer from fast solvent exchange, causing broadening of spectral peaks in indirect dimension. In contrast, the three HISQC

sequences which record the evolution of in-phase coherence (N_x) produce much sharper signals. Given this distinctive behavior, we have set up the experiments such as to obtain optimal performance in relation to their intrinsic properties.

For simplicity let us focus on the data in the right column, where the effect of fast solvent exchange is especially prominent. The SOFAST-HMQC and FHSQC data shown in the right column have been recorded with only 4 increments in indirect dimension ($t_1 = 8$ ms); because of the rapid loss of magnetization there is no need in lengthy evolution period. The line broadening is extreme and the spectral resolution is poor, making it impossible to identify two distinct peaks separated by 0.3 ppm in ^{15}N dimension. Due to excellent water preservation properties, $^1H^{\varepsilon}$ magnetization is quickly replenished in these experiments. As a consequence, the data can be recorded with high repetition rate, which makes the experiments highly sensitive (cf. panels K and L in Fig. 2).

On the other hand, all three HISQC sequences have been recorded with 64 increments ($t_1 = 175$ ms). The decay of N_x magnetization is slow, which allows for long t_1 evolution and translates into high resolution in ^{15}N dimension. If desired, the resolution can be further improved by extending t_1 period, thus taking advantage of the fact that intrinsic ^{15}N R_2 relaxation in peptides and disordered proteins is typically very slow at the physiological temperature of 37 °C, on the order of 1 s^{-1}. In principle, along these lines one can obtain spectra with very sharp ^{15}N resonances, reminiscent of small molecules. Recording such spectra would require somewhat longer experimental time, ca. 1–2 h.

While HISQC experiments clearly display superior resolution, their sensitivity is a source of concern. A substantial amount of signal is lost due to solvent exchange during the refocused INEPT periods. Furthermore, the water preservation properties are not as good as in SOFAST experiment, which necessitates the use of longer recycling delays. All of this translates into mediocre sensitivity. The panel M in Fig. 2 illustrates arginine $^{15}N^{\varepsilon}$-$^1H^{\varepsilon}$ spectrum recorded by means of the HISQC sequence (Iwahara et al. 2007). The two peaks are borderline detectable, with S/N ratios of 11 and 6. Certain improvement can be obtained by implementing a better water preservation scheme. Specifically, we have found that 3-9-19 WATERGATE (Sklenar et al. 1993) achieves better results than the original version employing water flip-back pulses. To quantitate this (trivial) improvement, we have set up the experiment which is different from the original HISQC only in that it uses 3-9-19 WATERGATE (HISQC-wg3919). The obtained moderate improvement can be judged by examining the data in panel N.

Of greater interest, however, are the gains associated with the use of CP element to transfer magnetization from ^{1}H to ^{15}N. Replacing the first refocused INEPT period in the pulse sequence with CP element results in twofold gain in peak intensity (cf. panels N and O in Fig. 2). The S/N ratios of the two observed peaks are increased to 31 and 16. Of note, CP-HISQC is the only experiment which obtains a fully satisfactory spectrum in ca. 10 min, allowing for reliable detection of the two well-resolved peaks.

Finally it is worth pointing out that original HISQC sequence has been developed for lysine side-chain NH_3 groups; in these four-spin systems the maximum attainable efficiency for both refocused INEPT and CP element does not exceed 44 %. Note, however, that CP-HISQC retains its advantage over the regular HISQC also in this case. The application to arginine $H^{\varepsilon}N^{\varepsilon}$ sites discussed above is free of this limitation, and so is the application to backbone H^NN sites, which is described in the next section.

CP-HISQC spectra of backbone amides

The N-terminal SH3 domain of the Drosophila adaptor protein drk (drkN SH3 domain) is a well-established model system for studies of protein disorder (Marsh and Forman-Kay 2009). The protein is marginally stable; in vitro it exists in a state of dynamic equilibrium between folded (F) and unfolded (U) forms. The exchange between the two states is slow, so that the spectra feature two distinct sets of peaks. From our perspective this is an attractive model for complex protein system, which combines the elements of order and disorder. Such complex systems increasingly become the focus of advanced NMR studies (Bista et al. 2012).

For our experimental measurements we have used the sample of drkN SH3 at pH 7.5, 30 °C. The proportion of F and U species under these conditions is approximately 1:1 (Zhang and Forman-Kay 1995). The unfolded form of drkN SH3 offers little protection against solvent exchange (Chevelkov et al. 2010). Therefore, it is appropriate to calculate solvent exchange rates using the simple algorithm which has been originally developed for random-coil peptides (Bai et al. 1993; Connelly et al. 1993). The results are dependent on protein primary sequence; for the unfolded drkN SH3, the majority of the predicted k_{ex} values fall in the range 15–150 s^{-1}. This is the regime where CP-HISQC is expected to be useful.[1]

In what follows we compare CP-HISQC with four other experiments that have been designed with fast solvent exchange in mind. Similar to the previous section, we have set up each experiment with consideration for its individual

[1] The alternative sample conditions could be pH 7.2, 37 °C. These conditions are relevant for cytosol (Orij et al. 2009). They give rise to very similar k_{ex} rates and are also suitable for CP-HISQC measurements.

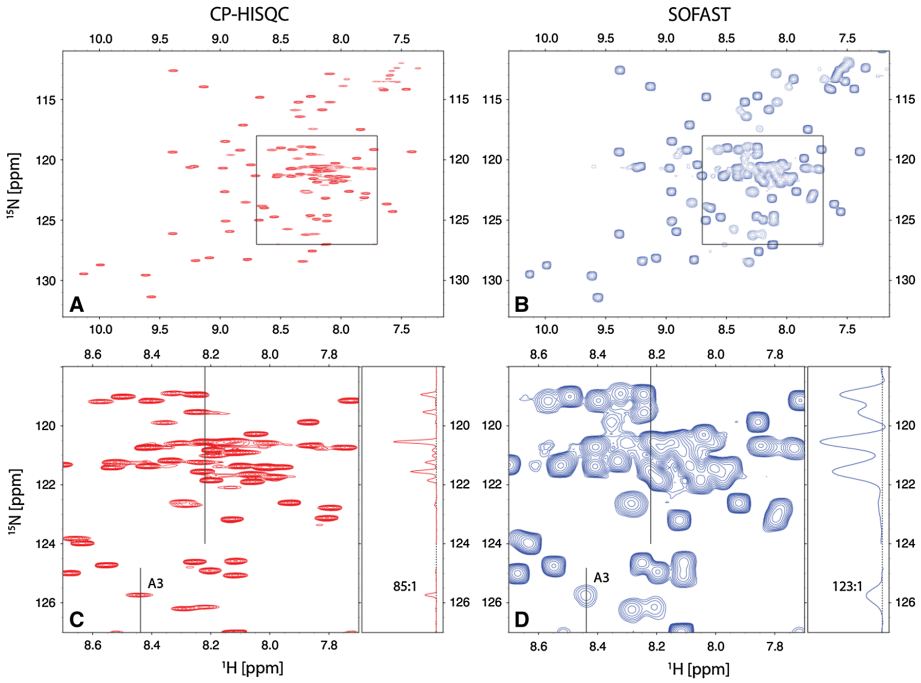


Fig. 3 (**a**, **b**) CP-HISQC and SOFAST-HMQC backbone amide spectra of drkN SH3. Each spectrum has been acquired in 50 min. The regions enclosed in rectangular boxes are magnified in panels (**c**, **d**). Black vertical lines indicate the position of 1D traces (displayed in *side panels*). The peak from residue A3 belonging to the unfolded form of drkN SH3 is labeled in the graph. For the purpose of plotting the spectra were scaled such that the noise level in all two-dimensional spectral maps is the same

properties. Specifically, SOFAST-HMQC and FHSQC have been recorded with 72 increments in indirect dimension, $t_1 = 39$ ms, and short recycling delays, $d_1 = 0.2$ s. At the same time, all HISQC experiments were recorded with 320 increments, $t_1 = 177$ ms, and standard recycling delay, $d_1 = 1.0$ s. Similar to the previous section, the best results have been obtained from SOFAST-HMQC (high sensitivity) and CP-HISQC (superior resolution in combination with reasonably good sensitivity). The corresponding spectra, both of which have been collected in ca. 50 min, are shown in Fig. 3.

The results in Fig. 3 demonstrate a dramatic improvement in resolution afforded by CP-HISQC experiment. The effect is especially obvious in panels C and D, which represent the most crowded portion of the spectrum. Indeed, a significant number of signals that are partially or fully overlapped in the SOFAST-HMQC spectrum turn out to be nicely resolved in CP-HISQC (cf. also 1D traces displayed in the side panels). This result is of major importance for studies of intrinsically disordered proteins or, more broadly, protein systems with elements of disorder. It is well known that the spectra of IDPs tend to be crowded and therefore high resolution is often key to the success of NMR studies. The increase in the number of quantifiable peaks should be very useful for various HSQC-based titration experiments. It also makes CP-HISQC a potentially useful template for ^{15}N R_2 experiment, PRE experiment, and other similar measurements.

In addition to superior spectral resolution, CP-HISQC experiment also offers reasonably good sensitivity. This can be appreciated by looking at A3 resonance which is labeled in the spectral map. The signal-to-noise ratio for this peak, which belongs to the U subspectrum, is 85:1 (see Fig. 3c, side panel). This is comparable to the S/N ratio found in the SOFAST-HMQC spectrum, 123:1 (Fig. 3d, side panel).

To further characterize the sensitivity of the two experiments, we have selected 22 peaks belonging to the U subspectrum which are well resolved and can be unambiguously assigned in all of the spectra (Zhang et al. 1994) (the assignment is shown in Fig. S1). For each of these spectral peaks we have determined the signal-to-noise ratio. The results are shown in Fig. 4a, where the x axis represents the intensity of CP-HISQC signals, while the y axis represents the *relative* intensity of CP-HISQC with respect to SOFAST-HMQC. On

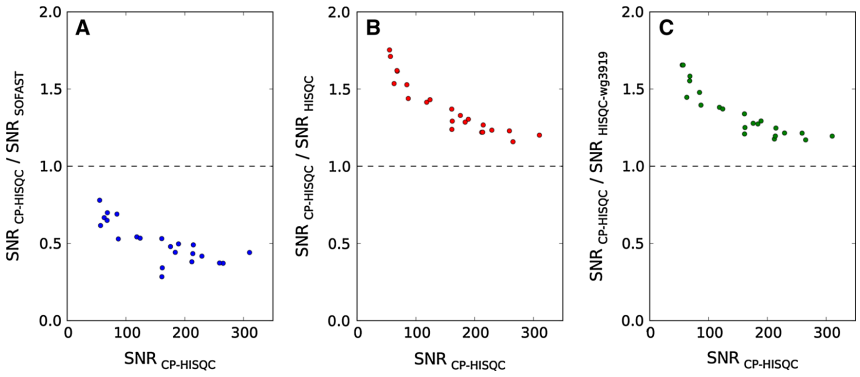


Fig. 4 Sensitivity of CP-HISQC experiment relative to SOFAST-HMQC, regular HISQC, and HISQC-wg3919 experiments (**a**, **b**, and **c**, respectively). Plotted along x axis: signal-to-noise ratio (SNR) of the individual peaks in the U subspectrum of CP-HISQC spectrum of drkN SH3. Plotted along y axis: relative signal-to-noise ratio of the same peaks, as obtained from comparison with other experiments. The SNR values are calculated using peak intensities (i.e. fitted maximum amplitudes), normalized by the noise level

average, the sensitivity of the CP-HISQC experiment proves to be 0.51 of that achieved by SOFAST-HMQC. It is worth noting, however, that CP-HISQC becomes increasingly competitive for the weaker signals (cf. the negative correlation in Fig. 4a). For the weakest peak, that from residue S10, the sensitivity of the CP-HISQC experiment relative to SOFAST-HMQC reaches 0.78.

A similar comparative analysis can be performed for the peaks from the F subspectrum. For the most part the F peaks are insensitive to exchange and therefore have much higher intensity than U peaks. Using the data from 44 well-resolved F resonances, we have found that the relative sensitivity of CP-HISQC compared to SOFAST-HMQC on average amounts to 0.36. In the case of surface residue N35, which experiences fast solvent exchange, this parameter rises to 0.57. In this connection it should be noted that CP-HISQC experiment has been designed to target U sites rather than well-protected F sites. At the same time, SOFAST-HMQC is known to be very efficient for F species, i.e. small globular proteins.

Of note, standard experiments such as SOFAST-HMQC cannot be easily optimized for simultaneously collecting data from U and F components at or near physiological conditions. Indeed, in U species the $2N_xH_x$ mode which is employed in SOFAST-HMQC undergoes rapid decay due to solvent exchange, $k_{ex} \sim 15–150\ s^{-1}$. On the other hand, in F species this mode is relatively long-lived, $R_{DQ/ZQ} \sim 1–5\ s^{-1}$ assuming the temperature 30–37 °C. This imposes conflicting requirements on the length of the t_1 evolution period. The new CP-HISQC experiment is free of this complication. Recall that CP-HISQC utilizes the N_x mode which decays slowly in both U and F species.

Of major interest to us is the contribution of the CP element. Figure 4b illustrates the sensitivity of CP-HISQC experiment relative to the previously reported HISQC experiment. For those U peaks that are well resolved and unambiguously assigned, the average increase in sensitivity proves to be 1.38. The most significant gain, 1.75, is again observed for the weakest peak, S10.

In part, these gains are due to the more efficient water preservation scheme implemented in CP-HISQC. To quantify this factor, we have implemented the reference experiment HISQC-wg3919. The only difference between CP-HISQC and HISQC-wg3919 is that the former starts with CP element, while the latter relies on refocused INEPT. The relative sensitivity of the two measurements is illustrated in Fig. 4c. The average gain from the use of the CP element is described by the factor 1.34.

Finally, CP element also works well for F species. All of the 44 selected F peaks are stronger in the CP-HISQC spectrum than in the HISQC-wg3919 spectrum. On average, the improvement is expressed by the factor 1.10. The biggest gain, 1.38, is again found in residue N35.

Based on all of the above, we suggest that CP-HISQC is the best choice of experiment for $^{1}H^{N}$–^{15}N spectroscopy of drkN SH3 at or near physiological conditions. The HISQC scheme achieves critical improvement in resolution, whereas the CP element offers significant boost in sensitivity.

Discussion

The efficiency of CP scheme: dependence on τ_{mix}

The advantage of CP-HISQC over more conventional schemes, such as HSQC, FHSQC, SOFAST-HMQC, stems from three factors:

(i) The reliance on the in-phase N_x evolution during the t_1 period. It is well appreciated that N_x has more favorable relaxation properties compared to $2N_xH_z$ or $2N_xH_x$ (Bax et al. 1990). In the context of this study it is especially important that N_x is immune to the effects of solvent exchange, whereas $2N_xH_z$ and $2N_xH_x$ suffer heavy losses (Iwahara et al. 2007).[2]

(ii) The ability of the CP element to efficiently utilize the influx of H_x magnetization from the H_2O proton pool, converting H_x into the desired N_x magnetization.

(iii) The more favorable properties of the CP element relative to the INEPT scheme with regard to spin relaxation (Levitt 1991) as well as solvent exchange. Both CP and refocused INEPT generate multispin modes on the route from H_x to N_x. However, the refocused INEPT has a higher content of multispin modes and, therefore, suffers greater losses due to solvent exchange.

The advantages associated with property (*i*) have been previously documented, although to the best of our knowledge the application to backbone amides undergoing fast solvent exchange is new. This aspect is also illustrated above (cf. the sharpening of spectral peaks in the HISQC family of spectra, Figs. 2 and 3). Therefore in what follows we focus on properties (*ii*) and (*iii*) wherein lies the novelty of our approach.

To analyze the details of CP transfer in the presence of fast solvent exchange we have conducted a series of numeric simulations. The results are presented in Figs. 5, 6, 7, 8 below. Briefly, we have simulated the behavior of the two-spin system, $^1H^N$–^{15}N, undergoing exchange with the large water proton pool. For simplicity, we have neglected (dipolar and CSA) relaxation. Other aspects, such as DIPSI-2 pulse train which is applied on-resonance or close to resonance with $^1H^N$ signal but off-resonance relative to the water line, are modeled in a fully realistic fashion. In Fig. 5 we analyze the efficiency of the CP transfer, i.e. the amount of the generated N_x magnetization, as a function of the mixing time τ_{mix}. Since the focus of this simulation is on the CP element, water preservation outside the CP element is of no interest at this point. Accordingly, we disregard the synchronization condition $\tau_{mix} = (n/2) \times T_{DIPSI2}$.

[2] In principle, loss of multispin correlations due to solvent exchange can also be described as "relaxation". However, we prefer to reserve this term for faster processes—specifically for the dynamic regime where $J_{NH}/k_{ex} \leq 1$ (including Redfield limit where $J_{NH}/k_{ex} \ll 1$ and the formula for scalar relaxation of the first kind applies (Abragam 1961)). Since in our study these conditions are not necessarily fulfilled we prefer a more general description referring to the "loss of multispin correlations due to solvent exchange" (Skrynnikov and Ernst 1999). In what follows the term "relaxation" refers to dipolar and CSA mechanisms.

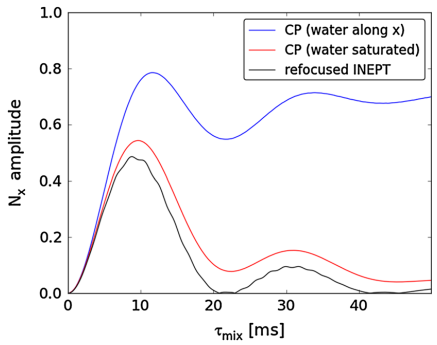


Fig. 5 Numerical simulations of proton-to-nitrogen magnetization transfer using CP or refocused INEPT element in the presence of solvent exchange with $k_{ex} = 100\ s^{-1}$. The simulations model the conditions and setup of the actual backbone experiment, including the use of DIPSI-2 sequence with $\nu_1^H = \nu_1^N = 4.0$ kHz and proton *rf* carrier positioned on-resonance with $^1H^N$ spin at 8.2 ppm (see "Materials and Methods" for additional details). Water magnetization is assumed to be initially aligned along x (*blue* and *black profiles*) or otherwise saturated (*red profile*). The refocused INEPT is assumed to be slightly asymmetric, $\tau_a/\tau_b = 0.9$, same as in the experimental measurement

Shown in Fig. 5 are three different profiles. The blue curve corresponds to the CP element as implemented in our experiment. Initially it is assumed that full amount of ^1H magnetization is available both on the amide and the water protons. Water ^1H magnetization is first aligned along the x axis and then effectively locked by the DIPSI-2 train. The resulting transfer profile is very different from the conventional $(1/2)(1–\cos(\pi J_{HN}\tau_{mix}))$ oscillatory pattern normally associated with the cross-polarization transfer. As can be seen from the plot, the transfer efficiency actually converges toward the plateau value of ca. 0.70. This behavior can be understood by noticing that fresh proton magnetization is constantly injected into the amide site via solvent exchange and then transformed into N_x. After sufficiently long τ_{mix} the system reaches a state of quasi equilibrium, $H_x \rightleftarrows 2(H_yN_z + H_zN_y) \rightleftarrows N_x$, where the loss of H_yN_z, H_zN_y due to solvent exchange is compensated by continuous influx of H_x from the large water pool.

The maximum transfer efficiency 0.79 is achieved with τ_{mix} of 11.6 ms. This is somewhat longer than the standard duration of the cross-polarization period, $\tau_{mix} = 1/J_{HN} =$ 10.8 ms. The additional time is required to utilize a portion of the incoming water polarization, maximizing the output of N_x. Note, however, that transfer efficiency remains very good in a broad range of τ_{mix}, from ca. 10 to ca. 13 ms. This provides us with the freedom to adjust τ_{mix} such as to satisfy the synchronization condition (see Fig. 1 and surrounding discussion). The distinctive dependence on τ_{mix}

as seen in Fig. 5 has been verified experimentally (shown in what follows).

The red profile in Fig. 5 describes the situation which is in every way similar, except it is assumed that water magnetization is saturated and remains zero throughout the course of the CP period. This scenario leads to a rather different outcome. Initially, amide magnetization H_x is transformed into N_x, with maximum transfer efficiency 0.54 reached at $\tau_{mix} = 9.6$ ms. As mixing time is increased beyond this level, amide $^1H^N$ magnetization is nullified due to the exchange with saturated water magnetization. Furthermore, the existing N_x magnetization is transformed back into H_x and subsequently destroyed through solvent exchange. As a result, the efficiency of the transfer drops to a very low level.

Finally, the black profile in Fig. 5 corresponds to the refocused INEPT transfer, where $\tau_{mix} = 2\tau_a + 2\tau_b$. In this case the behavior of water magnetization is different and deserves a special discussion. During the delays τ_a both amide and water ^1H magnetizations undergo free precession. Their precession frequencies are different by several thousand Hz. As a result, when water magnetization enters the amide site, it does so with a random phase (relative to the amide magnetization). Consequently, the input from water magnetization cancels out and the situation becomes generally similar to the previous scenario, where water signal is saturated.[3]

What are the main lessons that can be learned from Fig. 5? Clearly, the biggest advantage associated with CP sequence stems from its ability to utilize water magnetization (blue curve). If water signal is saturated, the performance of the CP scheme deteriorates (red curve). In fact, the efficiency of the CP element becomes comparable to that of the refocused INEPT (black curve). Hence it is mostly the water magnetization which makes the CP element successful, cf. factor (*ii*) in the above list.

At the same time we note that CP element is less susceptible to the exchange losses than refocused INEPT, corresponding to factor (*iii*) in the above discussion. This becomes apparent from the comparison of red and black profiles in Fig. 5. In both simulations there is no influx of magnetization from water, yet CP holds an advantage since it is less affected by the loss of $2H_yN_z$ and $2N_yH_z$. This advantage, however, is not as significant as the gain from water magnetization. Therefore we conclude that it is mainly factor (*ii*) which makes CP scheme superior, whereas factor (*iii*) plays a smaller role. This observation holds true for the entire range of k_{ex} rates relevant to our experiment.

[3] This general discussion is also valid in the situation when INEPT element includes pulsed field gradients.

What are the ramifications of the results in Fig. 5 for the pulse sequence Fig. 1 and possible alternative schemes?

First, it is worth noting that refocused INEPT element does not need to be symmetric. In fact, the first half of the refocused INEPT, $H_x \rightarrow 2H_yN_z$, is more vulnerable to solvent exchange than the second half, $2N_yH_z \rightarrow N_x$. Indeed, in the context of the INEPT scheme H_x, $2H_yN_z$, and $2N_yH_z$ are all effectively destroyed by solvent exchange, whereas N_x remains unaffected. Additional simulations establish that asymmetric INEPT using $\tau_a/\tau_b = 0.9$ shows a slightly better performance, with transfer efficiency increased by ca. 1 %. Generally, it is common to use somewhat different settings for τ_a and τ_b in the refocused INEPT, and we follow this path both in our experimental measurements and in numeric simulations.

Second, the conventional INEPT scheme can be modified such as to effectively lock water magnetization and utilize it toward generating N_x. Specifically, this can be accomplished by means of CPMG-INEPT (Muller et al. 1995; Mulder et al. 1996) with sufficiently high pulse repetition rate. We have simulated such hypothetical experiment and found that is indeed more sensitive than its conventional counterpart. However, the improvement proves to be limited, well below the level achieved by the CP scheme. This can be readily understood by noticing that CPMG-INEPT is useful only with regard to $H_x \rightarrow 2H_yN_z$ transfer, i.e. as first half of the refocused INEPT, but does not contribute anything during the second half. Therefore, we have not pursued this idea any further.

Third, we have considered the idea of a pulse sequence using two CP elements (for direct and reverse transfer). Relying on CP element to transfer magnetization from ^{15}N to ^1H has only limited advantages. Indeed, in this case water magnetization cannot be used to boost the signal. On the other hand, CP element remains less susceptible to exchange losses, cf. factor (*iii*) in the preceding discussion. In principle, this property offers ca. 10–15 % gain in observed signal (see Fig. 5) in addition to what has been achieved in the current CP-HISQC experiment. However, in practice the design using two CP elements is problematic. In particular, inserting CP element prior to the acquisition period makes it difficult to preserve water magnetization while at the same time minimizing residual water signal. Furthermore, second CP element increases the load on the probe, possibly requiring longer recycling delays. Considering these complications, we did not make any attempt to implement such experiment.

The efficiency of CP scheme: dependence on k_{ex}

Amide exchange rates in disordered proteins are dependent on primary sequence. As a result, an IDP sample presents a spectroscopist with relatively broad range of k_{ex} values. If

the sample also includes partially or fully folded components, as is the case with drkN SH3, then there is an additional set of signals with very low k_{ex}. Generally one may expect to encounter the situation where the peaks in the spectrum show a range of k_{ex} extending from 0 to $\sim 200\ s^{-1}$. Ideally, all of these peaks should be captured in a single experiment. As it turns out, the CP scheme can effectively deal with this situation: one τ_{mix} setting is sufficient to capture all resonances with near-optimal sensitivity. Specifically, the consensus setting $\tau_{mix} = 10.7$ ms achieves transfer efficiency within 6 % of the optimal level over the entire range of k_{ex} rates. The refocused INEPT scheme is somewhat less forgiving, i.e. a single τ_{mix} value cannot accommodate the range of k_{ex} values quite as well.

The range of k_{ex} rates from 0 to ca. 200 s^{-1} is generally suitable for $^1H^N$–^{15}N spectroscopy. In this region the advantage of the CP element over refocused INEPT is expressed by the average factor 1.5 (see Fig. 6). The area outside this region is more problematic from the perspective of $^1H^N$–^{15}N spectroscopy. For solvent exchange rates exceeding 500 s^{-1} refocused INEPT becomes completely inefficient, although CP element retains a reasonably good performance. In what follows we investigate the performance of CP element on the sample of Sos at elevated temperature where k_{ex} reaches 620 s^{-1}. It is demonstrated that even under these extreme conditions it is possible to obtain CP transfer with ca. 70 % efficiency (although this requires the use of longer mixing times).

We have also conducted spin dynamics simulations for Lys NH_3 groups. The results were found to be similar to those shown in Fig. 6, suggesting that CP element should be equally useful in the context of lysine side-chain experiments (Iwahara et al. 2007). As a caveat, note that the maximum transfer efficiency in this four-spin system cannot exceed 44 % (this limit holds for both refocused INEPT and CP schemes).

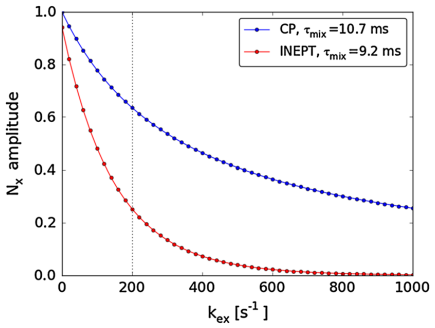


Fig. 6 The efficiency of CP element versus refocused INEPT transfer as a function of amide solvent exchange rate k_{ex}. The simulation parameters, including τ_{mix} settings, are the same as used in the actual experimental measurements (see "Materials and Methods")

The efficiency of CP scheme: dependence on *rf* field settings

Two conditions need to be met in order to achieve high efficiency of the CP element: (i) cross-polarization transfer per se should be effective, i.e. amide proton magnetization should be locked and fulfill Hartmann-Hahn condition, $\nu_1^H = \nu_1^N$, and (ii) water magnetization should be effectively locked along the same axis as amide proton magnetization. Both requirements could be readily satisfied if infinitely strong *rf* fields were available. In practice, however, ν_1^N is subject to rather stringent technical limitations, which automatically puts a limit on ν_1^H $(= \nu_1^N)$. As a consequence, ν_1^H is actually comparable to the offset Δ between the amide $^1H^N$ resonances and the water signal.

To address this aspect of the problem we have conducted a series of numeric simulations. In particular, the simulated data shown in Fig. 7a (red curve) imitate the conditions of our drkN SH3 experiment. In these simulations we assume that proton *rf* carrier is placed in the middle of the amide region, 8.2 ppm, such that the offset relative to that water signal is 3.5 ppm ($\Delta = 2.1$ kHz). We further assume that the strength of the DIPSI–2 *rf* field is $\nu_1^H = 4.0$ kHz and the solvent exchange rate is $k_{ex} = 100\ s^{-1}$. Using this set of parameters we have calculated the CP "excitation profile", i.e. determined the amount of N_x magnetization generated by the CP element for the range of $^1H^N$ chemical shifts. As it turns out, the entire amide region extending from 7.2 to 9.2 ppm is covered in a fairly uniform fashion (red curve in Fig. 7a). Somewhat higher efficiency is achieved at the upfield end of the spectrum compared to the downfield end, 0.80 vs. 0.72. This observation can be readily understood. Recall that CP transfer is efficient when amide and water magnetization are both locked along the x axis. This is easier to achieve for $^1H^N$ magnetization precessing at 7.2 ppm than for $^1H^N$ magnetization precessing at 9.2 ppm.

For the problem at hand the following relationship always holds true: $\nu_1^H, \Delta \gg k_{ex}$.Under this condition it can be shown that the efficiency of the CP element depends on the dimensionless ratio ν_1^H/Δ.[4] Figure 7b shows how the efficiency of CP transfer changes with the increase in ν_1^H/Δ. It is clearly desirable to use stronger *rf* fields in order to boost the sensitivity of the experiment. The inspection of the simulated data in Fig. 7b suggests that $\nu_1^H/\Delta \sim 2.0$ should be considered a reasonable target for CP-HISQC

[4] This parameter is the same as $\tan\theta$, where θ is the standard tilt angle.

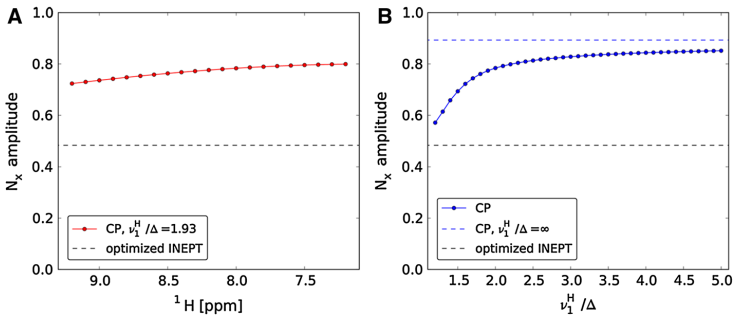


Fig. 7 The efficiency of CP element for backbone amides as a function of (**a**) $^1H^N$ chemical shift and (**b**) the strength of *rf* field used in CP mixing, $\nu_1^H(=\nu_1^N)$, relative to the offset between the proton carrier and the water signal, Δ. The simulations assume DIPSI-2 mixing with $\tau_{mix} = 10.7$ ms, proton *rf* carrier positioned at 8.2 ppm, and $k_{ex} = 100\ s^{-1}$. The results in (**b**) additionally assume that proton *rf* carrier is on-resonance with $^1H^N$ spin

measurements. Increasing the strength of *rf* field beyond this level produces only marginal improvement (cf. the curve in the graph which levels off and converges toward the plateau). In our experimental study of drkN SH3 the parameter ν_1^H/Δ was 1.93. The corresponding CP excitation profile is shown in Fig. 7a. In the study of the Sos peptide ν_1^H/Δ was 2.70 since $^1H^\varepsilon$ resonates closer to the water resonance and hence the offset Δ is smaller.

As already indicated, the limiting factor in setting up the CP-HISQC experiment is the nitrogen *rf* power. Excessive power leads to sample heating and may cause probe arcing (Keifer 1999) which severely degrades the performance of the pulse sequence. In the worst-case scenario, the probe can be permanently damaged. In our drkN SH3 measurements we have applied $\nu_1^N = \nu_1^H = 4.0$ kHz for 10.7 ms and found these settings to be safe. To quantify the amount of heating, we have prepared the sample where the coaxial insert tube was filled with methanol. This sample has been used to determine the temperature during the course of the CP-HISQC experiment as recently described (Yuwen and Skrynnikov 2013). The sample heating proved to be modest, 0.4 °C. Roughly half of the effect comes from the CP element, whereas the other half is due to proton decoupling during t_1. We have repeated these measurements on a sample additionally containing 100 mM NaCl and found that in this case the temperature increased by 0.8 °C. If necessary, one can compensate for such heating effects by adjusting the setting of variable-temperature unit.

It is worth noting that the problems associated with *rf* power are greatly alleviated in the latest-generation probes (Ban et al. 2012). The new probes can easily meet and surpass the $\nu_1^H/\Delta = 2.0$ condition. This is true not only for 500–600 MHz spectrometers, but also for higher field machines. The sample heating remains modest, even for high-salt samples, which can be attributed to the improvements in probe head design.[5] There is also an additional benefit from using stronger *rf* fields during the CP period as it helps to preserve water magnetization (see next section). All of this augurs well for future applications of the CP-HISQC experiment.

Next let us turn to the question of the proton carrier setting. From the perspective of Hartmann-Hahn transfer, it is advisable to set proton *rf* carrier in the middle of the amide region, at around 8.2 ppm. On the other hand, to simultaneously lock amide and water magnetization one may prefer to place the carrier between the two respective signals, at around 6.5 ppm. Which choice is better from the standpoint of the overall transfer efficiency? The simulations demonstrate that 8.2 ppm is very close to an optimal choice. A small improvement can be obtained by slightly shifting the carrier toward water. For $\nu_1^H/\Delta = 2.00$ this improvement becomes inconsequential (0.2 %). Therefore, we advocate 8.2 ppm, corresponding to the middle of the $^1H^N$ spectral region, as a standard choice for proton carrier setting.

Finally, let us touch upon the issue of pulse calibration. Significant miscalibration of the DIPSI-2 proton pulse reduces the efficiency of the Hartmann-Hahn transfer and, in addition, has an adverse effect on water preservation. Nevertheless, a small amount of error is well tolerated—deliberately mis-setting DIPSI-2 pulses by ±2 μs (3 %) causes virtually no change in the performance of CP-

[5] With this new hardware it is likely possible to execute the CP-HISQC experiment with shorter recycling delays. If in addition one can speed up the recovery of water magnetization, e.g. by introducing paramagnetic water relaxation agents (Hiller et al. 2005; Theillet et al. 2011), this can further improve the sensitivity of the CP-HISQC experiment. One should be mindful, however, that relaxation agents may perturb a delicate conformational equilibrium in the sample of an IDP.

HISQC. While it is recommendable to use HSQC-based sequence to calibrate proton pulses, it is also fine to use the standard water-based calibration procedure.

Preserving water magnetization

For labile protons in our drkN SH3 experiment solvent exchange rates mostly fall in the range 15–150 s^{-1}. On the other hand, the duration of the recycling delay d_1 is 1.0 s. Under these conditions, it is clear that amide $^1H^N$ magnetization is completely replaced with water magnetization during d_1 and therefore the optimal length of the recycling delay is dictated by recovery of water (rather than the relaxation properties of $^1H^N$). Consequently, it is important to prevent water saturation during the pulse sequence[6] and instead preserve water magnetization, such as to avoid long recovery periods.

CP-HISQC sequence is designed such as to ensure good water preservation properties. First, water magnetization is placed along the x axis by hard ^1H pulse.[7] Then it is effectively locked by a train of DIPSI-2 pulses, all of which have phase x or −x. After that water magnetization is returned to the z axis prior to application of the pulsed field gradient. Note that this scheme is subject to certain requirements regarding the synchronization of τ_{mix} with T_{DIPSI2} and the choice of ν_1^H, which have already been discussed above. The same tactics is used during the t_1 period, where the water is first placed along x, then effectively locked by the train of (on-resonance) WALTZ pulses with phase x (−x), and after that returned to z axis. Finally, during the refocused INEPT period, the effort is made to keep water magnetization along the z axis. The success of the water-preservation strategy can be conveniently characterized by the so-called water flip-back ratio f, which compares water magnetization before and after each scan, $f = M_{z,H2O}^{after}/M_{z,H2O}^{before}$ (Hiller et al. 2005).

We have determined the value of f for different experimental sequences used in this study (all measurements were conducted on the sample of Sos at 30 °C). In doing so we followed the procedure described by Hiller and co-workers. Considering 1D version of the experiments (i.e. t_1 set to zero) we have found that FHSQC and SOFAST-HMQC boast flip-back ratio f of 97 and 98 %, respectively. Given that these sequences preserve water magnetization so nicely, they can be executed with very short recycling delays. On the other hand, the HISQC-family sequences do not fare as well, with $f = 0.89$ found in the original HISQC, 0.93 in HISQC-wg3919, and 0.94 in CP-HISQC. These values typically require recycling delays on the order of 1 s. For $f = 0.94$ and $d_1 = 1$ s it is easy to estimate that CP-HISQC experiment reaches the steady state where ca. 75 % of the equilibrium water magnetization are available prior to each scan (Hiller et al. 2005). This result has been confirmed by additional experimental measurements (not shown).

Note that the coefficients f show substantial dependence on t_1. For instance, in the CP-HISQC experiment water magnetization experiences certain loss while being locked by WALTZ pulse train during t_1 period. For long t_1 evolution time, $t_1 = 175$ ms, this loss amounts to ca. 15 %. Therefore the f values derived from 1D experiments may not be ideally suited to determine the optimal recycling delays. As an alternative, we have recorded several trial 2D spectra and thus determined the d_1 values which ensure (near) optimal S/N ratio. Based on this standard procedure, the interscan delay was set to 0.2 s for SOFAST-HMQC and FHSQC experiments, while it was set to 1.0 s for HISQC, HISQC-wg3919, and CP-HISQC. All of these settings are consistent with the duty cycle requirements.

Finally, in all of our measurements we have used external lock, with D_2O added into coaxial insert tube (Wilmad, WGS-5BL). Conversely, adding D_2O directly to the protein sample in the presence of fast solvent exchange may cause a number of undesirable effects (Yuwen and Skrynnikov 2013).

CP transfer for amides with very fast solvent exchange

In the previous sections we have demonstrated that HSQC-style scheme can be used with the samples where solvent exchange rates k_{ex} approach 200 s^{-1}. Above this threshold, however, HSQC-type experiments become impractical because of extensive line broadening and loss of magnetization during the transfer steps. This is especially relevant for Lys and Arg side chains, where solvent exchange at pH 7.4, 37 °C is exceedingly fast. Many of the backbone $^1H^N$–^{15}N sites also suffer from very fast solvent exchange under physiological conditions.

In this situation one may opt for an alternative sequence design, different from a regular HSQC-like scheme. Specifically, let us consider a design where magnetization transfer starts from $^1H^N$, whereas detection is done on another spin, such as $^1H^\alpha$ or $^{13}C'$ (Bermel et al. 2009; Novacek et al. 2011; Gil et al. 2013). In what follows, we demonstrate that CP-based transfer $H_x \rightarrow N_x$ is well suited for use in any such sequence. In fact, the CP element

[6] Note that in the case of (fully or partially) folded proteins the saturation of the water signal does not only hurt labile $^1H^N$ protons, but also those protons that are protected from solvent exchange. Specifically, the saturation is transferred via spin diffusion from solvent-exposed protons on the surface of the protein to protons in the hydrophobic core, thus reducing the amount of magnetization across the board.

[7] Of interest, certain advanced schemes exist to align water magnetization along the effective *rf* field (Hansen and Kay 2007). These schemes, however, are only useful for spin-lock experiment and do not help in the case of phase-alternated pulse trains such as DIPSI-2.

proves to be indispensable if the spectra are to be recorded in the presence of very fast solvent exchange. Indeed, the simulations shown in Fig. 6 suggest that CP transfer remains fairly efficient for k_{ex} up to 1,000 s^{-1}. This is in stark contrast to INEPT which becomes totally ineffective in this regime.

In order to test these predictions experimentally, we have implemented a specialized pulse sequence for use on Arg side chains. Briefly, the magnetization transfer follows the path $H_x^\varepsilon \rightarrow N_x^\varepsilon(t_1) \rightarrow C_x^\delta \rightarrow H_x^\delta(t_2)$. The first transfer step is accomplished by means of the CP element which is identical to the one described in Fig. 1. The remaining steps are implemented in a standard fashion. The new sequence has been termed (HE)NE(CD)HD (see Fig. S2).

The (HE)NE(CD)HD spectra have been recorded on the sample of Sos peptide at pH 6.0. In addition to the data at 37 °C, we have also collected the data at 44 and 50 °C with the goal to emulate fast solvent exchange conditions observed under physiological pH. The exchange rates for Arg side-chain $H^\varepsilon N^\varepsilon$ sites were determined using the adapted version of the pulse sequence by (Segawa et al. 2008). Since the signals from four arginine residues are heavily overlapped, a single effective k_{ex} value has been obtained at each temperature: k_{ex} = 170, 330 and 620 s^{-1}, respectively.

Figure 8 shows the build-up of (HE)NE(CD)HD signal as a function of CP mixing time. In addition to the experimental data (circles), the plot also contains the simulated profiles (solid lines). These profiles have been calculated on the basis of the independently determined k_{ex} values, as described above. The only adjustment made in Fig. 8 involves the overall intensity scaling; other than that, no adjustable parameters have been used in generating this plot. We note that the level of agreement between the experimental and simulated data is highly satisfying.

As already pointed out, the shape of the N_x build-up profile in the presence of fast solvent exchange is different from the familiar oscillating curve. The transfer of magnetization from water to amine (or amide) proton and farther on to ^{15}N is controlled by solvent exchange, which is essentially a stochastic process. As a result, the CP build-up profiles acquire smooth character typical of stochastic processes (e.g. spin relaxation).

Note that in order to obtain efficient CP transfer, one needs to employ long τ_{mix}. Considering the data where k_{ex} = 620 s^{-1} (red curve in Fig. 8), the transfer efficiency ca. 50 % is achieved after 15 ms and the transfer efficiency ca. 70 % after 30 ms. This should be borne in mind if CP element is to be used in the presence of ultra-fast solvent exchange, $k_{ex} > 1{,}000$ s^{-1}. In this case the need for long τ_{mix} may constitute a technical limitation because of potential probe heating problem. Note also that in the case of long τ_{mix} one has to reckon with partial loss of water magnetization during the mixing period (in part due to saturation transfer from protein to water).

Also shown in Fig. 8 are the data from (HE)NE(CD)HD experiment where the starting CP element is replaced with

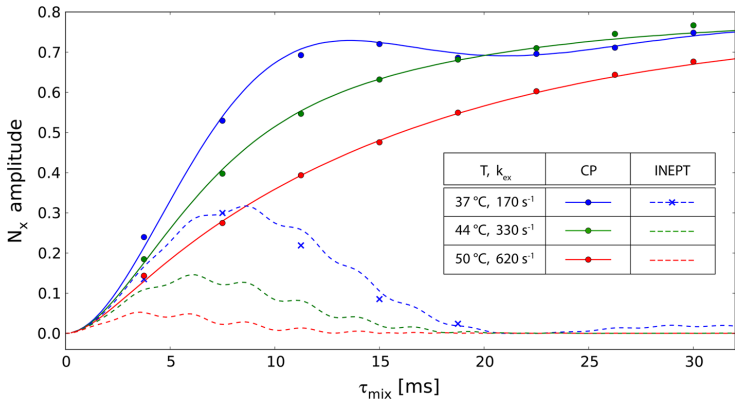


Fig. 8 The efficiency of CP transfer vs. INEPT transfer in a system with very fast solvent exchange (arginine side-chain $N^\varepsilon H^\varepsilon$ sites in Sos peptide). The experimental data have been collected using (HE)NE(CD)HD sequence employing proton-to-nitrogen CP element (*circles*) or, alternatively, refocused INEPT (*crosses*). The experiment was executed in 1D mode; the efficiency of the transfer has been estimated based on the intensity of the signal which represents four (overlapped) H^δ resonances. The solvent exchange rates k_{ex} listed in the legend have been measured independently using the suitably adapted version of the experiment by (Segawa et al. 2008). Also shown are the results of numeric simulations for CP and refocused INEPT transfer (continuous and dashed lines, respectively). To enable the comparison between the experiment and the simulations, the overall intensity scaling has been applied to the data. Specifically, for all data recorded at 37 °C, including both CP and INEPT data, the signal intensities have been multiplied by a constant scaling factor. This factor has been adjusted such as to match the simulated profile. The same procedure has been used for the data collected at two other temperatures

a regular refocused INEPT. The experimental data obtained at 37 °C are indicated by blue crosses; the simulated data pertaining to three different temperatures are shown as dashed curves. Clearly, the efficiency of INEPT transfer in the presence of very fast solvent exchange is low. For instance, in the case where $k_{ex} = 620\ s^{-1}$ the efficiency of the INEPT transfer is meager 5 % (red dashed curve in the plot Fig. 8). Obviously, it is not practical to use ^{1}H-to-^{15}N or ^{15}N-to-^{1}H INEPT transfer under very fast solvent exchange conditions.

Measurements targeting arginine, lysine, and histidine side chains, including side-chain sites undergoing very fast solvent exchange, play increasingly prominent role in protein NMR studies (Takayama et al. 2008; Trbovic et al. 2009; Blaum et al. 2010; Diehl et al. 2010; Esadze et al. 2011; Hu et al. 2012). The use of the CP element in this context is expected to be highly helpful.

Conclusion

Hartmann-Hahn cross-polarization in liquids is a highly efficient way of transferring magnetization. It has been shown that CP element has somewhat more favorable relaxation properties than refocused INEPT (Levitt 1991). CP transfer is also better protected against losses associated with conformational exchange (Krishnan and Rance 1995). Furthermore, CP element using DIPSI-2 significantly outperforms refocused INEPT in the presence of substantial *rf* field inhomogeneity (Majumdar and Zuiderweg 1995).

Fifteen years ago it has been shown that CP transfer is also more efficient than refocused INEPT for amide sites undergoing fast solvent exchange (Zangger and Armitage 1998). However, this observation was made in the context of a rather specialized experiment, exchange-edited HSQC. The pulse sequence developed for that purpose did not take advantage of in-phase (N_x) evolution in the indirect dimension. Also no special effort has been made to preserve water magnetization. Indeed, the role of water in the CP transfer has not been properly understood. As a consequence, the findings of that work have never found way into general-use HSQC spectroscopy. In our study we have addressed this omission by integrating the CP element into high-resolution HISQC experiment.

NMR studies of intrinsically disordered proteins are confronted with a number of significant challenges. The main difficulty arises from poor dispersion of chemical shifts in the spectra of IDPs. One should also bear in mind that the samples of IDPs often need to be prepared with low protein concentration because of the IDPs' propensity to aggregate and precipitate. Along the same lines, the measurement time is often limited because the samples tend to quickly deteriorate. If one attempts to conduct the measurements close to physiological conditions, the situation is further exacerbated by fast solvent exchange. Fast solvent exchange leads to intensity loss and dramatic line-broadening in the conventional HSQC-type spectra, making the spectroscopic studies very difficult or impossible.

In this paper we demonstrate how the complications associated with fast solvent exchange can be largely avoided and even turned into an advantage. Proton-decoupled experiment (HISQC) allows one to record the spectra with narrow peak width in ^{15}N dimension. Raising the temperature to (physiologically relevant) 37 °C achieves further sharpening of the signals in ^{15}N dimension, which effectively compensates for exchange-induced broadening in ^{1}H dimension. While the idea of proton-decoupled HSQC experiment (HISQC) is well-known, the benefits of this scheme in the context of backbone $^{1}H^{N}$–^{15}N spectroscopy of IDPs have not been previously appreciated.

The fact that water protons rapidly exchange into $^{1}H^{N}$ sites can also be used to one's advantage. First, as has been long realized, fast solvent exchange makes it possible to shorten the recycling delays. Second, as we found in this study, fast solvent exchange can be used to boost the efficiency of the CP transfer. The CP element is also a natural fit with the proton-decoupled HISQC scheme, thus leading to the CP-HISQC sequence described in this paper.

The use of the CP-HISQC scheme does not need to be limited to HSQC-style applications. This sequence can be used as a template for the experiments to measure paramagnetic relaxation enhancements (PREs), residual dipolar couplings (RDCs), etc. In particular, CP-HISQC sequence employs N_x coherence and thus it is perfectly suited for ^{15}N R_2 relaxation measurements. More broadly, the concepts used in CP-HISQC are also applicable to 3D sequences such as HNCO, HNCA, and others. Of interest, we have found that the properly designed CP element remains efficient even under very fast exchange conditions, $k_{ex} \sim 500$–$1000\ s^{-1}$. This observation suggests that proton-to-nitrogen CP transfer can be used as starting point in the pulse sequences that use non-labile ^{1}H or ^{13}C spins for detection. We have experimentally demonstrated this strategy for $H^{\varepsilon}N^{\varepsilon}$ sites in arginine side chains. Generally we believe that the pulse sequences and concepts presented in this work should significantly facilitate the spectroscopic characterization of disordered proteins at or near physiological conditions.

Materials and methods

Sample preparation

Sos peptide has been expressed in ^{13}C, ^{15}N enriched M9 media in a form of GB1-Sos fusion construct and purified by cation-exchange and gel-filtration FPLC. The peptide was

subsequently cleaved by application of formic acid (Vidovic et al. 2009) and purified by reverse phase HPLC. The concentration of the peptide has been determined by comparing the volume of V4 H^{γ} peak in 1D ^{1}H spectrum of Sos with the corresponding signal from 10 mM reference sample of valine. The conditions of the Sos sample were 1.0 mM Sos, 20 mM phosphate buffer in H_2O, pH 6.0, 0.02 % NaN_3. drkN SH3 has been expressed in ^{15}N enriched M9 media and purified according to the standard protocol (Zhang and Forman-Kay 1995). Protein concentration has been determined via UV absorbance at 280 nm. The sample conditions were 2.0 mM drkN SH3, 20 mM phosphate buffer in H_2O, pH 7.5, 0.02 % NaN_3. In both samples, Sos and drkN SH3, D_2O has been added into coaxial insert (Wilmad, WGS-5BL) to serve as external lock. This strategy avoids numerous complications that arise from use of H_2O/D_2O solvent under fast exchange conditions (Jurt and Zerbe 2012; Yuwen and Skrynnikov 2013).

NMR spectroscopy

All NMR measurements were performed on Varian Inova 600 MHz spectrometer equipped with a z-gradient TXI triple-resonance probe (after this study had been completed, CP-HISQC sequence was translated and successfully tested on Bruker DRX 500 MHz machine). 1D CP-HISQC experiment has been used to optimize *rf* field settings for the CP element. In doing so, we scanned ^{15}N power levels while maintaining ^{1}H power level fixed; we found it unnecessary to use fine power adjustments. The CP-HISQC arginine side-chain experiments on Sos have been recorded at 23, 30, and 37 °C (temperature calibrated using methanol). The spectra were collected with spectral width 6 ppm in ^{15}N dimension ($t_1 = 175$ ms) and 15 ppm in ^{1}H dimension ($t_2 = 46.5$, 23.3, and 11.8 ms for the three respective temperatures, chosen according to the formula $t_2 = 2/k_{ex}$ since the decay of the useful signal is dominated by solvent exchange, see Fig. 2). The data were apodized using phase-shifted sine-bell squared window function and zero-filled to the size 512 × 4,096. The CP-HISQC spectrum of drkN SH3 was recorded at 30 °C with spectral width 30 ppm in ^{15}N dimension and 15 ppm in ^{1}H dimension ($t_1 = 177$ ms, $t_2 = 40$ ms, chosen according to the formula $t_2 = 2/k_{ex}$ based on the compromise value $k_{ex} = 50\ s^{-1}$ for solvent exchange rate). The data were apodized and zero-filled to the size 2,560 × 4,096. The peak intensities have been obtained using nlinLS function in NMRPipe (Delaglio et al. 1995). The average noise level was estimated with the help of the function showApod. The solvent exchange rates were measured using the experiment by (Segawa et al. 2008) implemented with some minor modifications (e.g. the altered sequence evolves N_x coherence during the t_1 period instead of $2N_xC_z$).

Numerical simulations

The efficiency of proton-to-nitrogen CP and INEPT transfer in two-spin ^{1}H–^{15}N system has been simulated using in-house Maple and Python scripts. All coherent evolution mechanisms such as chemical shifts, *J*-couplings and *rf* pulses have been included in the simulations in a fully realistic fashion. The modeling of solvent exchange effect follows the scheme of (Skrynnikov and Ernst 1999), where all two-spin modes experience decay with the rate k_{ex}, whereas H_x^N mode exchanges magnetization with the corresponding water mode. The ratio of water magnetization to protein proton magnetization is assumed to be 1,000:1, reflecting protein concentration ca. 1 mM and assuming that a small-sized protein contains 100 exchangeable protons. The results do not change if one assumes the ratio 10,000:1, corresponding to ca. 0.1 mM sample. Given that spin relaxation in IDPs at or near physiological temperature is relatively slow and the mixing time τ_{mix} is short, we ignore dipolar and CSA spin relaxation. Likewise, we have neglected the potential effect of radiation damping noticing that water magnetization is either locked by strong *rf* field (CP) or dephased by application of pulsed field gradients (INEPT). The size of the Liouville superoperator matrix used in the simulations is 18 × 18 (comprising 15 modes from $^1H^N$–^{15}N spin system and 3 modes from water ^{1}H spin). The initial conditions have been chosen assuming that equilibrium magnetization is available for both $^1H^N$ and water. Alternatively, we considered somewhat artificial scenario where the water is initially saturated, but $^1H^N$ retains the full amount of magnetization (see Fig. 5).

Acknowledgments This study was supported by the funds from NSF Grant MCB 1158347.

References

Abragam A (1961) The principles of nuclear magnetism. Clarendon Press, Oxford

Amezcua CA, Harper SM, Rutter J, Gardner KH (2002) Structure and interactions of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation. Structure 10:1349–1361

Bai YW, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. Proteins Struct Funct Genet 17:75–86

Ban D, Gossert AD, Giller K, Becker S, Griesinger C, Lee D (2012) Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. J Magn Reson 221:1–4

Bax A, Ikura M, Kay LE, Torchia DA, Tschudin R (1990) Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. J Magn Reson 86:304–318

Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R (2006) ^{13}C-detected *protonless* NMR spectroscopy of proteins in solution. Prog NMR Spectrosc 48:25–45

Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P (2009) H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J Magn Reson 198:275–281

Bista M, Freund SM, Fersht AR (2012) Domain-domain interactions in full-length p53 and a specific DNA complex probed by methyl NMR spectroscopy. Proc Natl Acad Sci USA 109:15752–15756

Blaum BS, Deakin JA, Johansson CM, Herbert AP, Barlow PN, Lyon M, Uhrin D (2010) Lysine and arginine side chains in glycosaminoglycan-protein complexes investigated by NMR, cross-linking, and mass spectrometry: a case study of the Factor H–Heparin interaction. J Am Chem Soc 132:6374–6381

Chevelkov V, Xue Y, Rao DK, Forman-Kay JD, Skrynnikov NR (2010) $^{15}N^{H/D}$-SOLEXSY experiment for accurate measurement of amide solvent exchange rates. Application to denatured drkN SH3. J Biomol NMR 46:227–244

Connelly GP, Bai YW, Jeng MF, Englander SW (1993) Isotope effects in peptide group hydrogen exchange. Proteins Struct Funct Genet 17:87–92

Croke RL, Sallum CO, Watson E, Watt ED, Alexandrescu AT (2008) Hydrogen exchange of monomeric α-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in Escherichia coli. Protein Sci 17:1434–1445

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on unix pipes. J Biomol NMR 6:277–293

Diehl C, Engstrom O, Delaine T, Hakansson M, Genheden S, Modig K, Leffler H, Ryde U, Nilsson UJ, Akke M (2010) Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. J Am Chem Soc 132:14577–14589

Esadze A, Li DW, Wang TZ, Bruschweiler R, Iwahara J (2011) Dynamics of lysine side-chain amino groups in a protein studied by heteronuclear ^{1}H-^{15}N NMR spectroscopy. J Am Chem Soc 133:909–919

Farrow NA, Zhang OW, Forman-Kay JD, Kay LE (1994) A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. J Biomol NMR 4:727–734

Farrow NA, Zhang OW, Forman-Kay JD, Kay LE (1995) Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. Biochemistry 34:868–878

Felli IC, Brutscher B (2009) Recent advances in solution NMR: fast methods and heteronuclear direct detection. ChemPhysChem 10:1356–1368

Fernandez C, Hilty C, Wider G, Guntert P, Wuthrich K (2004) NMR structure of the integral membrane protein OmpX. J Mol Biol 336:1211–1221

Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC (2013) NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. Angew Chem Int Edit 52:11808–11812

Gray FLV, Murai MJ, Grembecka J, Cierpicki T (2012) Detection of disordered regions in globular proteins using ^{13}C-detected NMR. Protein Sci 21:1954–1960

Grzesiek S, Bax A (1993) The importance of not saturating H_2O in protein NMR: application to sensitivity enhancement and NOE measurements. J Am Chem Soc 115:12593–12594

Haddad KC, Sudmeier JL, Bachovchin DA, Bachovchin WW (2005) α-Lytic protease can exist in two separately stable conformations with different His^{57} mobilities and catalytic activities. Proc Natl Acad Sci USA 102:1006–1011

Hansen DF, Kay LE (2007) Improved magnetization alignment schemes for spin-lock relaxation experiments. J Biomol NMR 37:245–255

Hiller S, Wider G, Etezady-Esfarjani T, Horst R, Wuthrich K (2005) Managing the solvent water polarization to obtain improved NMR spectra of large molecular structures. J Biomol NMR 32:61–70

Hsu STD, Bertoncini CW, Dobson CM (2009) Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. J Am Chem Soc 131:7222–7223

Hu F, Schmidt-Rohr K, Hong M (2012) NMR detection of pH-dependent histidine-water proton exchange reveals the conduction mechanism of a transmembrane proton channel. J Am Chem Soc 134:3703–3713

Iwahara J, Jung YS, Clore GM (2007) Heteronuclear NMR spectroscopy for lysine NH3 groups in proteins: unique effect of water exchange on ^{15}N transverse relaxation. J Am Chem Soc 129:2971–2980

Jurt S, Zerbe O (2012) A study on the influence of fast amide exchange on the accuracy of ^{15}N relaxation rate constants. J Biomol NMR 54:389–400

Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114:10663–10665

Keifer PA (1999) 90° degrees pulse width calibrations: how to read a pulse width array. Concept Magnetic Res 11:165–180

Krishnan VV, Rance M (1995) Influence of chemical exchange among homonuclear spins in heteronuclear coherence-transfer experiments in liquids. J Magn Reson A 116:97–106

Kupce E, Freeman R (1996) Optimized adiabatic pulses for wideband spin inversion. J Magn Reson A 118:299–303

Levitt MH (1991) Heteronuclear cross polarization in liquid-state nuclear magnetic resonance: mismatch compensation and relaxation behavior. J Chem Phys 94:30–38

Loh SN, Kay MS, Baldwin RL (1995) Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. Proc Natl Acad Sci USA 92:5446–5450

Majumdar A, Zuiderweg ERP (1995) Efficiencies of double-resonance and triple-resonance J cross-polarization in multidimensional NMR. J Magn Reson A 113:19–31

Marion D, Ikura M, Tschudin R, Bax A (1989) Rapid recording of 2D NMR spectra without phase cycling: application to the study of hydrogen exchange in proteins. J Magn Reson 85:393–399

Marsh JA, Forman-Kay JD (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. J Mol Biol 391:359–374

Mori S, Abeygunawardana C, Johnson MO, van Zijl PCM (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. J Magn Reson B 108:94–98

Mueller GA, Smith AM, Chapman MD, Rule GS, Benjamin DC (2001) Hydrogen exchange nuclear magnetic resonance spectroscopy mapping of antibody epitopes on the house dust mite allergen Der p 2*. J Biol Chem 276:9359–9365

Mulder FAA, Spronk CAEM, Slijper M, Kaptein R, Boelens R (1996) Improved HSQC experiments for the observation of exchange broadened signals. J Biomol NMR 8:223–228

Muller L, Legault P, Pardi A (1995) Improved RNA structure determination by detection of NOE contacts to exchange-broadened amino protons. J Am Chem Soc 117:11043–11048

Novacek J, Zawadzka-Kazimierczuk A, Papouskova V, Zidek L, Sanderova H, Krasny L, Kozminski W, Sklenar V (2011) 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11

Orij R, Postmus J, Ter Beek A, Brul S, Smits GJ (2009) In vivo measurement of cytosolic and mitochondrial pH using a pH-

sensitive GFP derivative in Saccharomyces cerevisiae reveals a relation between intracellular pH and growth. Microbiology 155:268–278

Palmer AG, Cavanagh J, Wright PE, Rance M (1991) Sensitivty improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. J Magn Reson 93:151–170

Pasat G, Zintsmaster JS, Peng JW (2008) Direct ^{13}C detection for carbonyl relaxation studies of protein dynamics. J Magn Reson 193:226–232

Rucker SP, Shaka AJ (1989) Broad-band homonuclear cross-polarization in 2D NMR using DIPSI-2. Mol Phys 68:509–517

Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127:8014–8015

Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR 33:199–211

Segawa T, Kateb F, Duma L, Bodenhausen G, Pelupessy P (2008) Exchange rate constants of invisible protons in proteins determined by NMR spectroscopy. ChemBioChem 9:537–542

Shaka AJ, Keeler J, Frenkiel T, Freeman R (1983) An improved sequence for broad-band decoupling: WALTZ-16. J Magn Reson 52:335–338

Sklenar V, Piotto M, Leppik R, Saudek V (1993) Gradient-tailored water suppression for ^{1}H–^{15}N HSQC experiments optimized to retain full sensitivity. J Magn Reson A 102:241–245

Skrynnikov NR, Ernst RR (1999) Detection of intermolecular chemical exchange through decorrelation of two-spin order. J Magn Reson 137:276–280

Takayama Y, Castaneda CA, Chimenti M, Garcia-Moreno B, Iwahara J (2008) Direct evidence for deprotonation of a lysine side chain buried in the hydrophobic core of a protein. J Am Chem Soc 130:6714–6715

Theillet FX, Binolfi A, Liokatis S, Verzini S, Selenko P (2011) Paramagnetic relaxation enhancement to improve sensitivity of fast NMR methods: application to intrinsically disordered proteins. J Biomol NMR 51:487–495

Trbovic N, Cho JH, Abel R, Friesner RA, Rance M, Palmer AG (2009) Protein side-chain dynamics and residual conformational entropy. J Am Chem Soc 131:615–622

Vidovic V, Prongidi-Fix L, Bechinger B, Werten S (2009) Production and isotope labeling of antimicrobial peptides in Escherichia coli by means of a novel fusion partner that enables high-yield insoluble expression and fast purification. J Pept Sci 15:278–284

Werbeck ND, Kirkpatrick J, Hansen DF (2013) Probing arginine side chains and their dynamics with carbon-detected NMR spectroscopy: application to the 42 kDa human histone deacetylase 8 at high pH. Angew Chem Int Edit 52:3145–3147

Wu XD, Knudsen B, Feller SM, Zheng J, Sali A, Cowburn D, Hanafusa H, Kuriyan J (1995) Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. Structure 3:215–226

Yao SG, Hinds MG, Murphy JM, Norton RS (2011) Exchange enhanced sensitivity gain for solvent-exchangeable protons in 2D ^{1}H–^{15}N heteronuclear correlation spectra acquired with band-selective pulses. J Magn Reson 211:243–247

Yuwen T, Skrynnikov NR (2013) Proton-decoupled CPMG: a better experiment for measuring ^{15}N R_2 relaxation in disordered proteins. J Magn Reson [ePub ahead of print]

Zangger K, Armitage IM (1998) Sensitivity-enhanced detection of fast exchanging protons by an exchange-edited gradient HEHA-HA-HSQC experiment. J Magn Reson 135:70–75

Zhang O, Forman-Kay JD (1995) Structural characterization of a folded and an unfolded state of an SH3 domain in aqueous buffer and solution structure of this domain by NMR. Biochemistry 34:6784–6794

Zhang OW, Kay LE, Olivier JP, Forman-Kay JD (1994) Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed-field gradient NMR techniques. J Biomol NMR 4:845–858

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Role of Electrostatic Interactions in Binding of Peptides and Intrinsically Disordered Proteins to Their Folded Targets. 1. NMR and MD Characterization of the Complex between the c-Crk N-SH3 Domain and the Peptide Sos

Yi Xue,[†,∥] Tairan Yuwen,[†] Fangqiang Zhu,[‡] and Nikolai R. Skrynnikov*[,†,§]

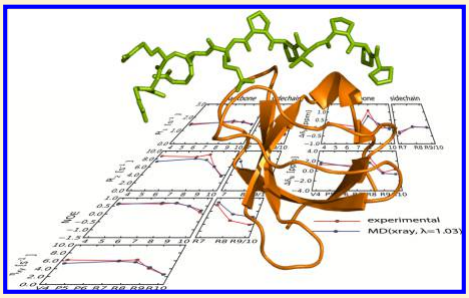
[†]Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States

[‡]Department of Physics, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana 46202, United States

[§]Laboratory of Biomolecular NMR, St. Petersburg State University, St. Petersburg 199034, Russia

Supporting Information

ABSTRACT: Intrinsically disordered proteins (IDPs) often rely on electrostatic interactions to bind their structured targets. To obtain insight into the mechanism of formation of the electrostatic encounter complex, we investigated the binding of the peptide Sos (PPPVPPRRRR), which serves as a minimal model for an IDP, to the c-Crk N-terminal SH3 domain. Initially, we measured ^{15}N relaxation rates at two magnetic field strengths and determined the binding shifts for the complex of Sos with wild-type SH3. We have also recorded a 3 μs molecular dynamics (MD) trajectory of this complex using the Amber ff99SB*-ILDN force field. The comparison of the experimental and simulated data shows that MD simulation consistently overestimates the strength of salt bridge interactions at the binding interface. The series of simulations using other advanced force fields also failed to produce any satisfactory results. To address this issue, we have devised an empirical correction to the Amber ff99SB*-ILDN force field whereby the Lennard-Jones equilibrium distance for the nitrogen−oxygen pair across the Arg-to-Asp and Arg-to-Glu salt bridges has been increased by 3%. Implementing this correction resulted in a good agreement between the simulations and the experiment. Adjusting the strength of salt bridge interactions removed a certain amount of strain contained in the original MD model, thus improving the binding of the hydrophobic N-terminal portion of the peptide. The arginine-rich C-terminal portion of the peptide, freed from the effect of the overstabilized salt bridges, was found to interconvert more rapidly between its multiple conformational states. The modified MD protocol has also been successfully used to simulate the entire binding process. In doing so, the peptide was initially placed high above the protein surface. It then arrived at the correct bound pose within ~2 Å of the crystallographic coordinates. This simulation allowed us to analyze the details of the dynamic binding intermediate, i.e., the electrostatic encounter complex. However, an experimental characterization of this transient, weakly populated state remains out of reach. To overcome this problem, we designed the double mutant of c-Crk N-SH3 in which mutations Y186L and W169F abrogate tight Sos binding and shift the equilibrium toward the intermediate state resembling the electrostatic encounter complex. The results of the combined NMR and MD study of this engineered system will be reported in the next part of this paper.



The interplay between order and disorder is an essential feature of any proteome. Those proteins in which disorder prevails are known as intrinsically disordered proteins (IDPs). They can be identified with relative ease on the basis of primary sequence; proteins with significant net charge and a low proportion of hydrophobic residues tend to be disordered.[1−3] Additional parameters, such as flexibility and bulkiness of amino acids, can also be incorporated into prediction algorithms.[4] Estimates obtained along these lines suggest that as many as 50% of eukaryotic proteins contain long stretches of disordered residues,[5] and more than 20% can be described as predominantly disordered.[4] These statistics demonstrate that IDPs cannot be dismissed as a rare quirk of nature; on the contrary, they represent one of the broad and fundamentally important classes of proteins.

Of interest, the proportion of IDPs in eukarya appears to be much higher than in archaea or prokarya (in the case of archaea, only 2% of proteins are predicted to have long disordered regions).[6] Thus, IDPs can be viewed as an advanced

Received: July 23, 2014
Revised: September 9, 2014
Published: September 10, 2014

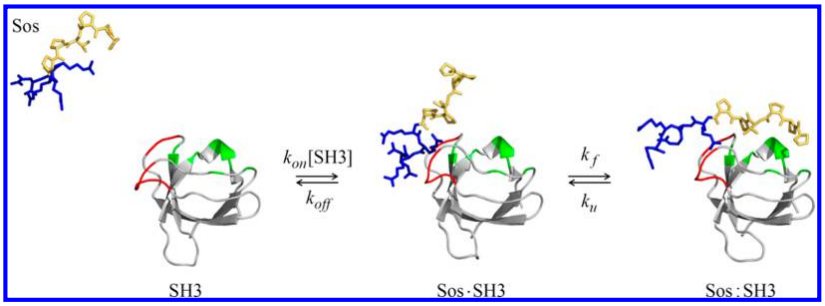


Figure 1. Schematic representation of the binding mechanism between the 10-residue peptide Sos and c-Crk N-SH3 domain. The images are from MD trajectories and are for the purpose of illustration only. The proteins are colored as follows: blue for positively charged Arg residues in Sos, red for negatively charged Glu and Asp residues that form an acidic patch at the periphery of the peptide binding site on the surface of the SH3 domain, gold for hydrophobic residues in Sos, and green for hydrophobic residues lining up the peptide binding site on the surface of the SH3 domain. Note that in reality free Sos and Sos·SH3 are mixtures of highly diverse conformational species.

type of molecular machinery that evolved in higher organisms. As it happens, disordered proteins are typically involved in cell signaling and regulation.[7,8] In this context, the unfoldedness of IDPs confers a number of functional advantages. IDPs can literally wrap themselves around their binding targets, thus making use of multiple binding sites.[9−13] This mechanism creates new possibilities in terms of binding affinity and specificity. For example, an intrinsically flexible protein may be able to bind several structurally different partners. Such multitasking helps in the development of efficient, highly integrated signaling networks, in which every protein serves in more than one role. In addition, IDPs offer the benefit of economy because they present large binding interfaces but do not require large supporting scaffolds.[14]

What are the functional requirements for an IDP binding to its structured target in the context of cell signaling? First, sufficiently high specificity is to be expected of any such interaction. This implies a fairly extensive pattern of contacts between an IDP and its folded target. To form these contacts, IDP needs to adopt an appropriate spatial conformation. Given that intrinsically disordered proteins normally go through a multitude of different conformations, this requirement should effectively slow the binding process and lead to low on rates (k_{on}). At the same time, there is an evolutionary pressure to maintain a high off rate (k_{off}). This is necessary to ensure a rapid response to changing external stimuli. The combination of a low k_{on} and a high k_{off} is expected to produce a large dissociation constant ($K_d = k_{off}/k_{on}$). In a cellular milieu, where the concentrations of proteins are typically low, this should render the entire mechanism unsustainable.

As it appears, nature has found a way to resolve this conundrum. As already pointed out, IDPs usually carry a significant net charge (which often includes a contribution from post-translational modifications, such as phosphorylation).[15−17] This fundamental characteristic of IDPs makes it possible to rely on the so-called electrostatic steering mechanism to boost the on rate.[18] Briefly, long-range electrostatic forces pull an IDP toward the relevant acidic (or basic) patch on the surface of a target protein, thus establishing contact in the vicinity of the binding site. The resulting state, in which IDP is loosely anchored at the periphery of the binding site, has been described as "electrostatic encounter complex".[19] This mechanism effectively increases the local concentration of the ligand in the vicinity of the binding site and thus promotes binding. The on rates in excess of 10^7 M^{-1} s^{-1} are invariably associated with this mechanism.[20,21]

After the electrostatic encounter complex is formed, it quickly evolves into the final complex, which is structured and has a (reasonably) high affinity. During this transition, the IDP adopts a suitable conformation to make multiple binding contacts. This process has been described as "folding upon binding".[22] (In general, binding of IDPs to their folded target falls under the rubric of "induced fit", although the elements of "conformational selection" have also been noted.[23,24]) It has also been observed that the individual segments of the IDP "coalesce" around the multiple attachment points on the surface of the target protein.[25,26] The electrostatic interactions continue to play a significant role at this late stage in the binding process. It has been pointed out that electrostatic complementarity is essential for binding specificity in the case of IDP complexes.[27] It has also been suggested that the folding is greatly accelerated by a pattern of charged residues surrounding the binding site.[28,29]

In this paper and the following paper, we focus on the details of the electrostatic interactions involved in the formation of the electrostatic encounter complex. We use the experimental nuclear magnetic resonance (NMR) data as well as molecular dynamics (MD) simulations to obtain a realistic model for these interactions. Recent advances in computer hardware and MD algorithms made it possible to simulate binding of a small IDP to its folded target using regular (unbiased) MD simulations in explicit solvent. This has been demonstrated for binding of peptides to modular domains such as SH2 and SH3 (where the peptides can be viewed as a minimal model for an IDP).[30,31] At the same time, Ubbink's laboratory recently presented an NMR study of low-affinity electrostatic complexes formed by short peptides with the sequence XKKKK and electron transfer protein plastocyanin.[32] In our work, we pursue a synergistic approach, seeking to combine experimental NMR data with MD simulations. Of note, we have used a rigorous MD protocol with a state-of-the-art force field, as opposed to coarse-grained simulations, replica exchange simulations, or Monte Carlo modeling, which were all successfully applied in the context of peptide docking.[29,32−34] The use of bona fide MD allows us to draw a direct comparison with the NMR measurables.

For the purpose of this study, we have selected a system consisting of the N-terminal SH3 domain from adaptor protein

c-Crk and a 10-residue peptide from Ras activator protein Sos. Cellular Crk is ubiquitously expressed in a wide range of cells and tissues.[35,36] In its role as an adaptor protein, c-Crk appears at the center of a vast and complex signaling network; the function of c-Crk is to bring together various proteins that bind to its several modular domains (SH2 and one or two SH3 domains).[37−39] In particular, the N-terminal SH3 domain of c-Crk binds a number of guanine nucleotide exchange factors; these interactions have been directly linked to the high motility of cancerous cells.[40−42] One of the nucleotide exchange factors that binds to c-Crk N-SH3 is the Ras activator protein Sos.[43−45] The proline-rich sequence that is responsible for this binding belongs to the disordered carboxyl-terminal segment of Sos.[46] The corresponding 10-residue peptide, PPPVPPRRRR, has been crystallized in complex with murine c-Crk N-SH3.[47] The resulting crystallographic structure was used in the recent MD simulation study featuring a shorter version of the Sos peptide, PPPVPPRR.[30]

In the first part of our study (this paper, hereafter termed paper 1), we focus on the interaction between the wild-type murine c-Crk SH3 (wtSH3) and the 10-residue Sos peptide (Sos). The presumed binding mechanism is illustrated in Figure 1, which shows the transition from the free state of the peptide (Sos) to the putative electrostatic encounter complex (Sos·wtSH3) and further to the fully formed complex (Sos:wtSH3). The formation of the hypothetical encounter complex is underpinned by the interaction between four arginines at the C-terminus of the Sos peptide (blue) and the Glu/Asp-rich patch on the surface of wtSH3 (red). The final complex is then formed as the side chains of the two conserved Sos prolines (gold) are tucked into the shallow hydrophobic grooves on the surface of wtSH3 (green). As a matter of convention, the formation of Sos·wtSH3 is classified as binding (rate constants k_{on} and k_{off}) and the subsequent transition to Sos:wtSH3 is viewed as folding (rate constants k_f and k_u).

For the Sos:wtSH3 complex, which has a relatively high affinity, we have experimentally measured backbone and side-chain ^{15}N relaxation rates in the bound Sos as well as spectral frequency shifts caused by the binding. We have also recorded a long MD trajectory of Sos:wtSH3 in explicit solvent under the Amber ff99SB*-ILDN force field,[48−50] as well as another long MD trajectory representing the free Sos peptide. The MD data have been used to predict NMR measurables, and the calculated results were subsequently compared with the experimental data. This comparison revealed a systematic problem with the MD simulation—as it turned out, the salt bridge interactions in Sos:wtSH3 were consistently over-emphasized. This is a known problem with the most popular force fields.[51,52] To address this problem, we have introduced an ad hoc correction to the force field. Specifically, the parameters of the pairwise Lennard-Jones potential between the ionized groups in the Arg and Glu/Asp side chains have been slightly adjusted. Using this modified version of the force field, we were able to successfully reproduce the experimental data from the Sos:wtSH3 complex. Furthermore, we were able to simulate the entire binding process beginning from free Sos, progressing to the Sos·wtSH3 encounter complex, and ultimately undergoing the transition to the final Sos:wtSH3 complex. This MD trajectory was also consistent with the experimental NMR data in the part pertaining to the Sos:wtSH3 state.

As a next step, we focused on the investigation of the electrostatic encounter complex between Sos and c-Crk N-SH3 (second part of this work, termed paper 2, to be submitted for publication). The experimental characterization of encounter complexes of IDPs poses a serious challenge. First, they tend to be sparsely populated and short-lived. Second, they are comprised of many interconverting conformational species. Third, they are typically found in rapid exchange with the heavily populated free and/or bound states. NMR spectroscopy, which is the most important experimental technique in this arena, is affected by all of the complications mentioned above. In principle, encounter complexes involving IDPs can be targeted using those experiments that are highly sensitive to the presence of minor species: relaxation dispersion,[53,54] chemical exchange saturation transfer,[55,56] or paramagnetic relaxation enhancements.[57] Yet it remains extremely difficult to capture the three-state exchange process such as that illustrated in Figure 1. For instance, a recent relaxation dispersion study of the Fyn SH3 domain binding 12-residue peptide failed to produce any direct evidence of the encounter complex.[58]

In this situation, we resort to an alternative strategy. Specifically, we have modified the interaction of Sos with c-Crk N-SH3 such as to mimic the electrostatic encounter complex. Toward this goal, we introduced two point mutations in the area of two shallow hydrophobic grooves in the SH3 domain (green ribbon in Figure 1). These two mutations, Y186L and W169F, are supposed to abrogate the hydrophobic component of binding that is responsible for formation of the final complex. At the same time, these mutations preserve the electrostatic Arg-to-(Asp/Glu) interactions. The idea is to shift the thermodynamic balance away from Sos:SH3 and toward the Sos·SH3 electrostatic encounter complex.

The altered system is designed to boost the population of the electrostatic encounter complex. This opens the door for standard NMR experiments, which do not need to be sensitive to weakly populated species. In particular, in the case of the electrostatic encounter complex between Sos and the double mutant SH3 domain (dmSH3), we have recorded HSQC titration data reporting on k_{on} and k_{off} rates, backbone and side-chain ^{15}N relaxation rates reporting on the mobility of the Sos peptide, the corresponding ^{1}H and ^{15}N chemical shifts, and the temperature coefficients pertaining to these chemical shifts. Using our adaptation of the Amber force field, we were able to produce an MD model that proved to be in good agreement with the experimental data from the Sos·dmSH3 system. All of the results pertaining to Sos·dmSH3 will be discussed in paper 2.

MATERIALS AND METHODS

Protein and Peptide Expression. A primer for producing c-Crk N-SH3 was designed in house using mouse cDNA (GenBank accession number BC031149) and GST plasmid pGEX-4T-1. The protein was expressed and purified as described elsewhere[59] to produce unlabeled, ^{15}N-labeled, or ^{15}N- and ^{13}C-labeled material. The ^{15}N-labeled Sos peptide was bacterially expressed as a fusion construct with a solubility tag protein, the GB1 domain from protein G.[54,60,61] The Sos sequence was subcloned into GB1-containing plasmid GEV 2, with a single spacer residue (Asp). Following the carefully optimized expression and purification procedure, GB1-Sos was cleaved at the Asp-Pro site using a treatment with a 50% (v/v) solution of formic acid (24 h at 50 °C). The products were separated using C18 column chromatography. The sample was concentrated using a lyophilizer or SpeedVac apparatus; the buffer exchange was conducted over a period of at least 24 h

dx.doi.org/10.1021/bi500904f | Biochemistry 2014, 53, 6473−6495

using a dialyzer device with a 0.5 kDa membrane cutoff. The yield from this procedure was 1−2 mg of ^{15}N-labeled Sos from 1 L of M9 medium. Unlabeled Sos, as well as Sos selectively labeled with ^{15}N at the V4 position, was purchased commercially (Peptide 2.0 Inc.).

Sample Conditions. Unless indicated otherwise, the samples used in this work have been prepared in a buffer containing a 90% H_2O/10% D_2O buffer, 20 mM phosphate, and 0.02% NaN_3 (pH 6.0). The concentrations of the peptide and protein were determined using the NMR method by Wider and Dreier, which ensures an accuracy on the order of 3%.[62,63] Specifically, we determined the intensities of isolated methyl resonances from one-dimensional (1D) proton spectra (V4 in Sos and V184 in c-Crk N-SH3) and compared them with the external reference (10 mM sample of valine). For c-Crk N-SH3, the results were consistent with the standard approach based on UV absorption by aromatic residues, A_{280}. All NMR data were collected at 23.4 °C.

NMR Measurements. NMR measurements were conducted using a Varian Inova 600 MHz spectrometer equipped with an HCN probe; additionally, ^{15}N relaxation parameters were measured using a Bruker Avance DRX 500 MHz spectrometer equipped with a TXI cryoprobe. The backbone ^{15}N relaxation measurements were conducted using updated versions of the standard relaxation experiments,[64−66] including the recently corrected heteronuclear NOE sequence.[67] ^{15}N CSA-dipolar transverse cross-correlation rate η_{xy} has been measured by means of the IPAP sequence[68] with an added relaxation period with a length that is a multiple of $1/^1J_{NH}$.[69] For arginine side-chain $^{15}N^{\varepsilon}$ relaxation measurements, 1D versions of the same experiments were used, with the nitrogen rf carrier set to 80 ppm. The data have been processed using the nmrPipe suite of programs,[70] including the *autofit* script, and Sparky.[71]

Spectral Assignments. The resonance assignments of wtSH3 are from ref 72. The resonance assignment for backbone amides in Sos is obtained from standard NOESY and TOCSY experiments performed on a 10 mM sample of unlabeled Sos.[73] The assignment has been confirmed by solvent exchange data: at pH 7.0, the resonances from R8 and R9 are broadened beyond detection, which is consistent with expectations.[74] They were also confirmed using a triple-resonance experiment correlating $^1H^{\varepsilon}$,$^{15}N^{\varepsilon}$ spins with $^1H^N$,^{15}N spins of the subsequent residue (unpublished). The spectrum of the free peptide also contains a number of minor peaks arising from the *cis* conformation of adjacent proline residues;[75] some of these minor peaks have also been assigned. Four arginine $^1H^{\varepsilon}-^{15}N^{\varepsilon}$ correlations in the spectrum of the free Sos peptide are overlapped, resulting in a single unresolved peak. Two of these resonances become resolved and can be assigned in the complex with wtSH3. $^1H^{\eta}-^{15}N^{\eta}$ correlations were found to be severely broadened and have not been used.

K_d and k_{on} for Binding of Sos to wtSH3. Isothermal calorimetry measurements were conducted at 23.4 °C in NMR buffer using an iTC200 instrument (MicroCal). The solution of wtSH3 (concentration 0.1 mM, volume 300 μL) was titrated with Sos (concentration 10 mM, injection volume 2 μL, total of 18 injections). The data were fit using MicroCal Origin, assuming two-state exchange:

$$\text{Sos} + \text{wtSH3} \underset{k_{OFF}}{\overset{k_{ON}[\text{wtSH3}]}{\rightleftharpoons}} \text{Sos:wtSH3} \quad (1)$$

which is the reduction of the three-state scheme shown in Figure 1. (The notations k_{ON} and k_{OFF} are reserved specifically for the two-state scheme in eq 1; they are distinct from k_{on} and k_{off} used in the three-state scheme in Figure 1.) The fitting suggests single-site binding with a K_d of 1.5 ± 0.1 μM (see Figure 2).

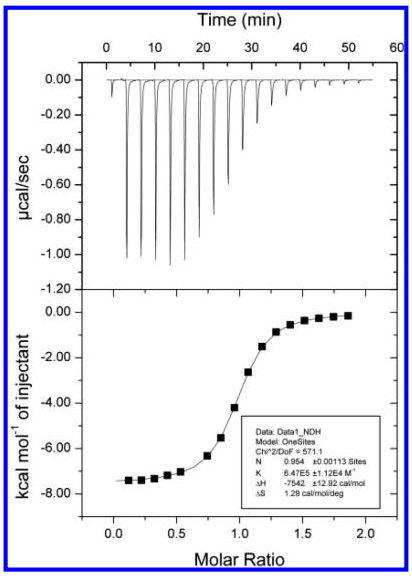


Figure 2. Calorimetric titration of wt c-Crk N-SH3 with Sos. The obtained affinity constant is 1.5 μM.

NMR titration was conducted by adding unlabeled wtSH3 to the sample of ^{15}N-labeled Sos (see Figure 3A). Large titration shifts are observed for residues V4, R7, and R8; for all these sites, the on−off exchange falls in the intermediate regime. The line shape analysis of V4 titration yields k_{ON} values of 1.6×10^9 and 2.0×10^9 M^{-1} s^{-1} (^{15}N and ^{1}H dimensions, respectively; see Figure 3B). At the same time, the joint treatment of all titration data from V4, R7, and R8 produces the global k_{ON} value of 1.5×10^9 M^{-1} s^{-1}.

Tumbling Time of wtSH3. The rotational correlation time of the protein needs to be known to complement MD simulations and facilitate the calculation of relaxation rates. For free wtSH3 in an aqueous solution, the calculation using HYDRONMR[79] predicts a nearly isotropic diffusion tensor with a τ_{rot} of 3.91 ns and a $D_{\parallel}/D_{\perp}$ of 1.05 (see Table S1 of the Supporting Information for details). To experimentally determine the parameters of overall tumbling, we relied on the ^{15}N R_2/R_1 data as analyzed by the program *r2r1_diffusion*.[80] The residues included in the analyses were selected according to the prescription of Tjandra et al.[81] Using the ^{15}N relaxation data from a dilute sample of free wtSH3 (protein concentration of 0.2 mM) and assuming that diffusion tensor is axially symmetric, we obtained τ_{rot} = 4.08 ns and a $D_{\parallel}/D_{\perp}$ = 1.14, in close agreement with expectations. However, as we increase the concentration of the protein to 2 mM, the experimentally determined τ_{rot} value shoots up to 6.64 ns. At the same time, the tumbling anisotropy remains nearly

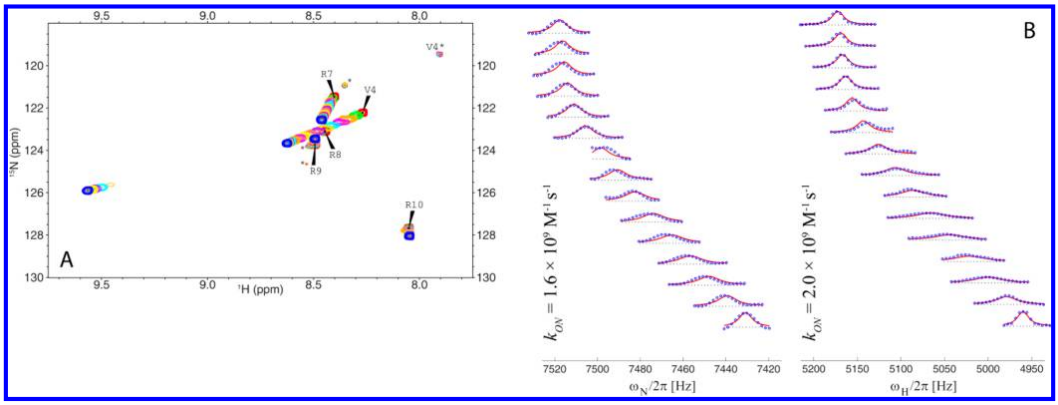


Figure 3. (A) $^{1}H^{N}-^{15}N$ HSQC titration of ^{15}N-labeled Sos with unlabeled wtSH3. The starting point corresponds to 0.20 mM Sos without wtSH3 (red contour in the plot); the final point corresponds to 0.18 mM Sos, after accounting for dilution, and 0.35 mM wtSH3 (blue contour in the plot). During the course of the titration, the peak from residue R8 disappears and then reappears far downfield. The minor peaks, labeled with asterisks, represent distinct conformational species arising from *cis*–*trans* isomerization of proline residues in Sos;[76,77] one of them has been identified as belonging to the valine residue (V4*). Minor peaks do not shift upon titration because the alternative conformational species Sos* have a much lower affinity for wtSH3. At the same time, these peaks gradually disappear as the protein:peptide ratio approaches 1:1. Indeed, with the addition of the protein, the Sos* ⇄ Sos ⇄ Sos:wtSH3 equilibrium becomes shifted to the right, thus depleting the population of the minor conformational species Sos*. The titration of the arginine side-chain N^{ε} sites has been recorded separately using the appropriate ^{15}N carrier setting (not shown). (B) Titration of spectral peak V4 (nitrogen and proton dimensions). The traces from HSQC spectra were fit using LineShapeKin.[78] Briefly, this program uses a fixed value of K_d as obtained from the ITC measurements; the line widths and intrinsic resonance frequencies are determined from the spectra of the free and protein-saturated Sos (bottom and top traces, respectively). The data from nitrogen and proton dimensions have been fit separately; the resulting k_{ON} values are indicated in the plot. The code of LineShapeKin has been slightly modified to obtain the desired graphics.

unchanged ($D_{\parallel}/D_{\perp}$ = 1.18), and the orientation of the long axis of the diffusion tensor changes only insignificantly, by 16° (Table S1 of the Supporting Information). This kind of behavior points toward low-affinity nonspecific (or weakly specific) dimerization, as has been previously observed for a variety of small globular proteins.[82−86]

Weak Dimerization of wtSH3. To quantitatively characterize the self-association effect in wtSH3, we have conducted a $^{1}H^{N}-^{15}N$ HSQC titration over the range of protein concentrations from 0.2 to 2.0 mM. A large number of signals are seen to titrate, shifting by a small amount (<0.06 and <0.3 ppm in the proton and nitrogen dimensions, respectively). The data have been fit assuming that the protein undergoes fast exchange between the monomeric and dimeric forms. The results of the fitting are presented in Figure S1 of the Supporting Information. Briefly, a dimerization site with an affinity constant in the range of 3−8 mM is found in the RT loop. In addition, there are multiple other sites with affinities of >10 mM that are broadly distributed over the surface of the protein. This is consistent with the notion that at a concentration of 2 mM roughly half of the protein is in the dimeric form, as can be deduced from ^{15}N relaxation data. The dimerization appears to be only weakly specific, which also agrees with the ^{15}N relaxation analyses (see above).

Tumbling Time of Sos:wtSH3. The τ_{rot} of the wtSH3:Sos complex has been determined using the same sample conditions that were used in the main series of relaxation measurements: 2 mM SH3 and 0.5 mM Sos. Addition of the peptide resulted only in a small increase in τ_{rot}, ca. 2% (see Table S1 of the Supporting Information). Considering that only one-fourth of the protein molecules are loaded with the peptide under the current experimental conditions, we estimate that the τ_{rot} of the Sos:wtSH3 complex is 7.2 ns (assuming an isotropic tumbling model).

Importantly, this result lends itself to direct experimental verification. The sample used to collect R_2/R_1 data was prepared with both wtSH3 and Sos peptide labeled with ^{15}N. Several of the Sos resonances are well-resolved in the spectra, including those from V4 and R8 that are in rigid contact with the protein (see the crystallographic structure Protein Data Bank entry 1CKB[47]). The R_2/R_1 ratio for these two residues translates into a tumbling time of 6.8 ns for the Sos:wtSH3 complex. More sophisticated model-free analysis using data collected at two magnetic field strengths yields values of 7.1 and 6.8 ns for the two peptide residues at hand. We conclude that the τ_{rot} value of 7.0 ± 0.2 ns represents an accurate and reliable result for the Sos:wtSH3 complex. For the purpose of subsequent MD analyses, it is reasonable to assume that the tumbling is isotropic; the modest anisotropy effect can be safely neglected.

MD Simulations. The main series of MD simulations has been conducted using the Amber 11 package with the ff99SB force field,[48,87] including ILDN side-chain correction[50] and backbone helical propensity correction.[49] The initial coordinates of the system were generated using two different approaches.

In the first approach, the initial coordinates were based on the crystallographic structure 1CKB.[47] The missing Sos residues R9 and R10 were added with random conformation using the program MODELLER.[88] The SH3 domain was simulated without three terminal residues, G132, S133, and R191, which were present in our NMR sample. We also conducted control simulations, in which these three residues were added to the crystallographic structure; the results proved to be unchanged.

dx.doi.org/10.1021/bi500904f | *Biochemistry* 2014, 53, 6473−6495

In the second approach, the peptide was placed over the surface of the protein in a random fashion. Specifically, in the crystal structure 1CKB, we defined a vector connecting the centers of mass of wtSH3 and Sos. The length of this vector was doubled from 12.6 to 25.2 Å, and the obtained point was used to place the center of mass of the Sos peptide. The initial conformation of peptide was chosen at random from a 2.1 μs MD simulation of the free Sos. The initial orientation of Sos was assigned randomly. Finally, a random rotation was applied to wtSH3 (the amplitude of this rotation was normally distributed with a variance of 30°).

All Glu and Asp residues in the SH3 domain were assumed to be deprotonated under the conditions of our experimental study (pH 6.0). This is consistent with the calculation employing the program PROPKA[89] and the crystallographic structure 1CKB. These calculated pK_a values are all sufficiently low (the highest value, 5.0, is predicted for residue E149). To verify this result, we have prepared a sample of Sos:wtSH3 with an elevated pH (7.0) where it is safe to assume that all Asp and Glu side chains are deprotonated. We have found that the chemical shifts of Sos, including $^{1}H^{\eta}$ and $^{15}N^{\eta}$ shifts, remain essentially unchanged in this sample. Therefore, we can conclude that Asp and Glu residues at the Sos binding interface are already deprotonated at pH 6.0.

The protein−peptide system thus generated was solvated using the explicit TIP3P solvent. A truncated octahedral water box was constructed such that the boundary of the water box was at least 12 Å away from any of the peptide or protein atoms. Given that the peptide is initially placed at considerable distance from the protein, the resulting water box contains a generous amount of water. As a result, the effective concentration of the protein in our simulations was relatively modest, 6−8 mM. This is not too far from the experimental NMR concentration. (The above description pertains to the setup with random placement of Sos. In the case when the initial configuration was based on 1CKB, we used the water box with a 14 Å margin. Because the 1CKB complex is sufficiently compact, the water box proved to be smaller, corresponding to an effective protein concentration 13−15 mM.)

The hydrated system was neutralized by adding one Na^+ ion (for the protein missing R191) and subjected to energy minimization (1000 steps using harmonic restraints with a force constant of 500 kcal mol^{-1} Å^{-2}, followed by 1000 steps with no restraints), then heated from 0 to 296 K, and equilibrated for 1 ns at 296 K. The production-stage MD simulations were conducted at 296 K using the NPT ensemble. During the simulations, all bonds involving hydrogen atoms were constrained using the SHAKE algorithm. The nonbonded cutoff was set to 8 Å. The integration step was 2 fs, and the coordinates were stored every 1 ps. The simulations were conducted using two GPU workstations, one equipped with four NVIDIA GeForce GTX480 cards and the other with four GTX580 cards (assembled by Electronics Nexus, Binghamton, NY, and Colfax International, Sunnyvale, CA). The production rate using the CUDA version of the pmemd program was in the range from 30 to 60 ns per day per card (depending on the size of the water box).

To adjust the strength of the salt bridge interactions in the Amber ff99SB*-ILDN force field, we have implemented one change in the original force field. The Lennard-Jones energy for a pair of atoms separated by distance r is parametrized in Amber as follows:

$$E_{L,J} = \varepsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right] \quad (2)$$

where ε is the well depth and r_0 is the equilibrium distance. We have altered the r_0 pertaining only to certain specific atom pairs, namely the nitrogen−oxygen pairs involving N^{η} atoms in Arg side chains and O^{ϵ}/O^{δ} atoms in Glu/Asp side chains. The original Amber-generated value of r_0 for these atomic pairs has been multiplied by the scaling factor $\lambda = 1.03$. Because the GPU version of the Amber 11 pmemd program does not use look-up tables, this was accomplished by making a simple modification to the source code (kNLCPNE.h).

The three main MD trajectories discussed in this paper are as follows:

- (*i*) The 3.06 μs trajectory of Sos:wtSH3 starting from the crystallographic coordinates 1CKB recorded with the standard force field, $\lambda = 1.00$. This trajectory is termed MD(xray,$\lambda = 1.00$).
- (*ii*) The 3.20 μs trajectory of Sos:wtSH3 starting from the crystallographic coordinates 1CKB recorded with the altered force field, $\lambda = 1.03$. This trajectory is termed MD(xray,$\lambda = 1.03$).
- (*iii*) The 2.28 μs trajectory that represents the binding of Sos to wtSH3 starting with the random configuration as described above. This trajectory has been recorded using the altered force field, $\lambda = 1.03$, and is termed MD(rand, $\lambda = 1.03$).

In addition, to quantitate NMR resonance shifts that occur in Sos upon binding to wtSH3, we have also recorded the 2.1 μs trajectory of the free Sos (because the peptide contains no Asp or Glu residues, there is no difference between the original and altered force fields in this case). Other MD trajectories recorded as a part of this study are discussed where appropriate in the text.

Processing of the MD Trajectories and Prediction of NMR Parameters. The MD trajectories were manipulated using in-house C++ package trjtool. This package supports both text and binary formats for coordinate files and various output files, includes a convenient batch file processing mode, and offers a range of functions (extracting atomic coordinates and torsional angles, superimposing structures, evaluating correlation functions and order parameters, etc.). Some of the functionalities have been adapted for use with the GPU. The companion package pytrj, programmed in python, serves to calculate and plot various relaxation parameters (including ^{15}N relaxation rates and PREs). Both packages are available upon request.

To calculate chemical shifts, we have extracted every 10th frame from a given trajectory, processed these frames using SHIFTX2 version 1.07,[90] and then averaged the results. All computations were conducted using the SHIFTX+ module, which deals with the conformational dependence of chemical shifts. To investigate the effect of sampling, we repeated the calculations using all MD frames; the results were found to be virtually identical. Our focus is on the differential chemical shifts, $\Delta\delta_{Sos} = \delta_{Sos}^{bound} - \delta_{Sos}^{free}$, which are termed binding shifts. In the calculations, the δ_{Sos}^{bound} component has been obtained from Sos:wtSH3 trajectories whereas δ_{Sos}^{free} has been obtained from the simulation of the free Sos. The effect of weak self-association of wtSH3 on $\Delta\delta_{Sos}$ is inconsequential and has been ignored in these calculations.

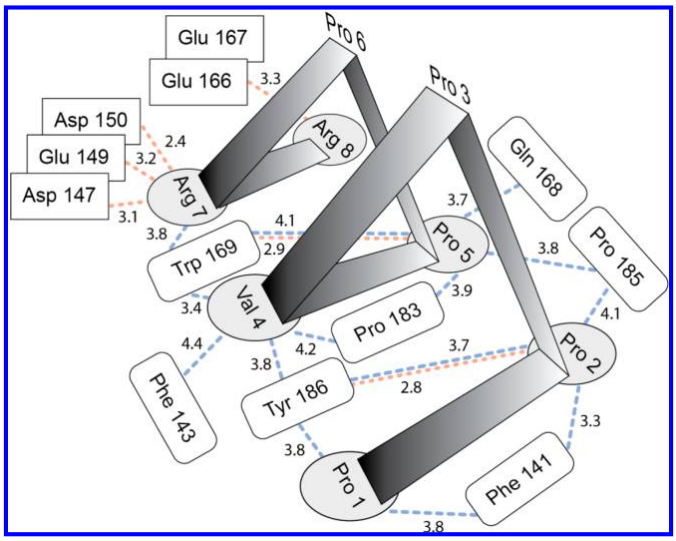


Figure 4. Binding of the Sos peptide to wild-type c-Crk N-SH3 as seen in the X-ray structure 1CKB.[47] The intermolecular distances indicated in the plot correspond to the nearest pairs of carbon atoms from hydrophobic residues (blue dashed lines) and nitrogen−oxygen pairs in the salt bridges or hydrogen bonds (pink dotted lines).

To calculate the relaxation rates, all MD frames were aligned with the crystallographic coordinates 1CKB by superimposing the protein coordinates (those of C^{α} atoms from the secondary structure residues). Note that only the protein coordinates have been superimposed; the motion of the peptide relative to the protein surface is fully preserved. The overall tumbling of the Sos:wtSH3 complex was then reintroduced using the experimentally determined correlation time, τ_{rot} = 7.0 ns (this value accounts for the effect of weak self-association as observed for wtSH3, see above). The autocorrelation functions have been computed for NH vectors using a sparse nonuniform time grid as described previously;[91] they were subsequently fit with six-exponential functions, multiplied by $\exp(-\tau/\tau_{rot})$, and then converted into spectral densities and further into relaxation rates.[92] In the case of CSA-dipolar cross-correlated cross-relaxation rate η_{xy}, the result was multiplied by $P_2(\cos\theta)$ assuming that the long axis of the ^{15}N CSA tensor makes an angle θ = 20° with the NH bond.[93] The length of the NH bond was assumed to be 1.02 Å, and the magnitude of the nitrogen chemical shift anisotropy was taken to be −172 ppm.

The in-house python script has been employed to analyze the appearance of salt bridges and hydrogen bonds in the MD trajectory. In the case of salt bridges, at least one nitrogen atom from the Arg side chain has been required to fall within 4 Å of the oxygen from the Glu/Asp side chain. In addition, the distance between the centroids of the two charged groups should not exceed 5 Å. This definition covers both high- and medium-strength ion−ion contacts, which mainly play a stabilizing role.[94] In the case of hydrogen bonds, we have required that the nitrogen−oxygen distance be <3.2 Å and the deviation from linearity not exceed 30°.[95]

Finally, the solvent-accessible surface area has been calculated on a per frame basis using the program POPS with the default parameter settings.[96]

RESULTS AND DISCUSSION

Overview of the Sos:wtSH3 Complex. The crystal structure of the complex between c-Crk N-SH3 and the Sos-derived peptide, PPPVPPRRRR, was determined by Wu and co-workers.[47] A schematic diagram of the binding interface is shown in Figure 4. The peptide is a class II ligand with consensus sequence PxhPx+ (x is any amino acid, h is a hydrophobic amino acid, and + is arginine or lysine).[97,98] Over most of its length, the peptide adopts a left-handed polyproline type II helix conformation. The helix has a triangular cross section; residues at the base of the triangle interact with the SH3 domain. The side chains of the two conserved prolines, P2 and P5, are tucked into the shallow hydrophobic grooves on the surface of SH3 (buried surface areas of 487 and 377 Å^2 for the peptide and protein, respectively). The grooves are defined by the side chains of conserved hydrophobic residues in the SH3 domain (represented by rounded boxes in Figure 4).

The positively charged C-terminus of the Sos peptide is delocalized and, therefore, partially absent from the crystallographic structure (no electron density for R10, R9, and the side chain of R8). The side chain of R7 is observed in the electron density map, even though crystallographic refinement reveals a substantial amount of variability at this site; the R7 guanidinium group forms salt bridges with carboxyl groups from D150, D147, and E149 (represented by rectangular boxes in Figure 4). One may expect that the side chain of R8 similarly interacts with E166 and E167,[47] although it must be significantly impacted by conformational disorder. Recall that electrostatic interactions involving the stretch of Arg residues in the Sos peptide anchor the putative encounter complex Sos·wtSH3. Given the dynamic nature of these interactions, this complex cannot be productively studied by X-ray crystallography. This situation calls for the development of new NMR-based methods.

dx.doi.org/10.1021/bi500904f | *Biochemistry* 2014, 53, 6473−6495

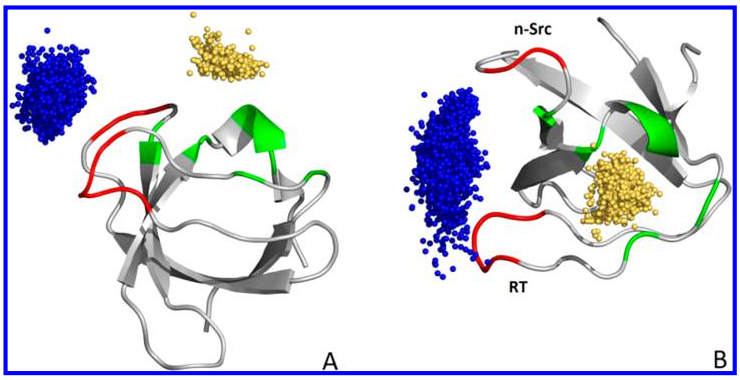


Figure 5. Position of the peptide Sos on the surface of wtSH3 according to the data from the 3.06 μs simulation MD(xray,λ = 1.00). To generate this plot, MD frames were first aligned with 1CKB by superimposing the SH3 coordinates (one frame per nanosecond). The peptide was then represented by the center of mass of its N-terminus (residues P1−P6, gold spheres) and C-terminus (residues R7−R10, blue spheres). In wtSH3, the charged residues at the binding interface are colored red (D147, E148, E149, and D150 in the RT loop and E166 and E167 in the n-Src loop) and the hydrophobic residues are colored green (F141, F143, W169, P183, P185, and Y186). Shown are the side view (A) and the top view (B) of the complex.

The dissociation constant (K_d = 5.2 ± 0.2 μM) has been previously reported for the binding of Sos to wtSH3 in the presence of 150 mM NaCl.[99] We were able to reproduce this result using the intrinsic fluorescence of residue W169 that resides at the ligand binding interface and thus is sensitive to Sos binding. Under identical conditions, our fluorescence titration data (not shown) yield a dissociation constant of 5.9 ± 0.2 μM. On the other hand, the ITC data collected under low-salt conditions, as also used in our NMR measurements, yield a K_d of 1.5 ± 0.1 μM (see Materials and Methods). As expected, the decrease in the ionic strength of the solution leads to stronger electrostatic attraction and, therefore, tighter binding.[100] The affinity of 0.1−10 μM is typical for binding of proline-rich peptides to SH3 domains and, more generally, for binding of peptides to their protein targets.[99,101] Such moderately strong binding is consistent with the signaling function that requires, on the one hand, a meaningful level of complex formation and, on the other, a rapid release of ligand in response to changing stimuli.

We have also analyzed the peak shapes from the NMR titration of the ^{15}N-labeled Sos with unlabeled wtSH3 (see Materials and Methods). The global fitting of the titration data using the fixed value of K_d as obtained in the ITC measurement allowed us to determine the on rate (k_{ON} = 1.5 × 10^9 M^{-1} s^{-1}). The off rate is then simply $k_{OFF} = K_d k_{ON}$. The approach in which ITC data are combined with NMR data to obtain a complete kinetic characterization of the system is similar to that employed by Demers and Mittermaier.[58]

The extremely high on rate obtained in our experiments is consistent with the prominent role of electrostatic interactions in bringing together Sos and wtSH3.[100,102−104] In fact, proteins that associate with rates >10^7 M^{-1} s^{-1} always seem to utilize the electrostatic steering mechanism. Previously, k_{ON} values of up to 5 × 10^9 M^{-1} s^{-1} have been measured for binding of barnase mutants to barstar.[100] An on rate as high as 1.7 × 10^{10} M^{-1} s^{-1} was recently found for the complex of the AD2 fragment from the intrinsically disordered N-terminal transactivation domain of tumor suppressor p53 with the zinc finger 2 domain of the CREB-binding protein.[105] These are extreme cases in which the proteins are propelled toward each other by strong electrostatic interactions.

Significant progress has been made in predicting k_{ON} based on atomic coordinates. In particular, the program TransComp produces results that are accurate to within 1 order of magnitude.[20] Although this program has been originally designed for folded proteins, it can also generate reasonable predictions for IDPs.[26] We used the selected MD frames as input for the TransComp calculations. Specifically, we have chosen the frames from the MD trajectory that represents the approach of Sos to wtSH3 and subsequent binding (see below). The k_{ON} rate estimated in this manner, ∼0.5 × 10^9 M^{-1} s^{-1}, is in good agreement with the experimental result.

The binding affinity is important from the perspective of NMR characterization. Our ^{15}N relaxation measurements and chemical shift determinations have been conducted on a sample containing 0.5 mM ^{15}N-labeled Sos and 2.0 mM unlabeled wtSH3. It is easy to verify that under these conditions 99.9% of Sos is bound. Therefore, the experimental NMR data pertain solely to the bound peptide.

Simulation of Sos:wtSH3 Using Amber ff99SB*-ILDN. Comparison with Experimental Data. The trajectory of Sos:wtSH3 has been recorded using the Amber 11 package with the ff99SB force field.[48,87] A number of recent comparative studies, in particular those based on the experimental NMR data, favor this force field over others.[106−111] In addition, recent corrections for selected backbone[49] and side-chain[50] torsional potentials have been employed. The revised force field is known as ff99SB*-ILDN.[112] The initial coordinates for the MD simulation have been taken from the crystallographic structure 1CKB. The C-terminal residues in Sos that are missing from the crystallographic model have been rebuilt using MODELLER.[88] The system was hydrated using TIP3P water.[113] Given that the C-terminal portion of Sos is dynamic and occasionally becomes lifted off the surface of wtSH3, we opted for a large water box (minimum 14 Å separation between any of the peptide or protein atoms and the water box boundaries). The resulting MD trajectory has a length 3.06 μs. In what follows, this simulation is termed MD(xray,λ = 1.00). The label indicates that the initial coordinates have been taken from the

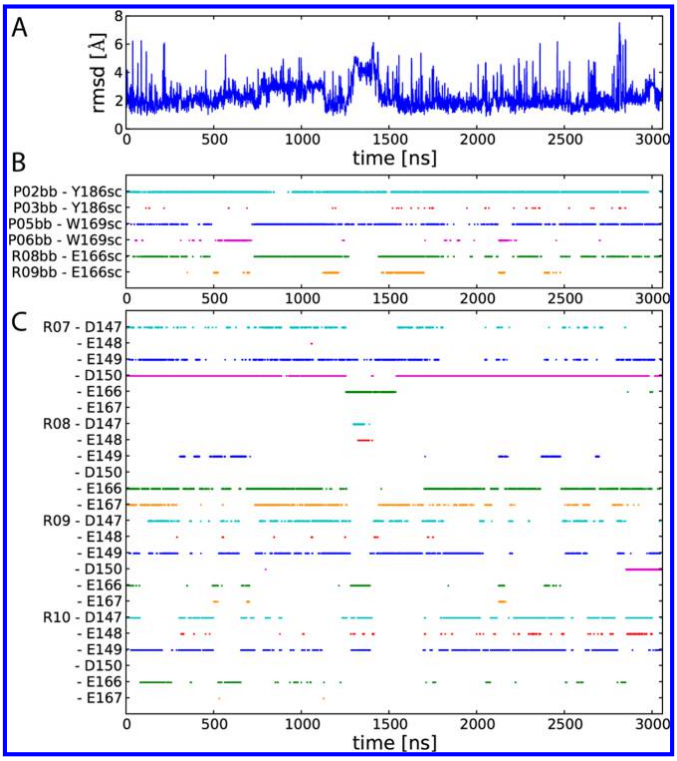


Figure 6. (A) rmsd of the peptide atomic coordinates relative to the crystallographic structure 1CKB as calculated for a 3.06 μs simulation of Sos:wtSH3 MD(xray,λ = 1.00). To generate this plot, MD frames were first aligned with 1CKB by superimposing the SH3 coordinates, and then the rmsd was computed for Sos atoms. Included in the rmsd calculation were all of the Sos atoms for which the crystallographic coordinates are available. (B) Peptide−protein hydrogen bonds and (C) salt bridges in the same Sos:wtSH3 trajectory (see Materials and Methods for definitions). The following procedure was used to generate these plots. The trajectory has been divided into 1 ns intervals, and the content of hydrogen bonds (salt bridges) has been determined for each interval. If the hydrogen bond (salt bridge) is present for more than half of the time during the 1 ns interval, then the dot is placed in the corresponding position in the graph. In generating panel B, we included only those hydrogen bonds that originate at the backbone of Sos and are populated at the level of at least 3%. In generating panel C, we have taken into consideration all possible combinations between the four Arg residues from the Sos peptide and the six Glu/Asp residues from the complementary acidic patch on the surface of the wtSH3 domain.

X-ray structure and the force field parameters have been used as is without any modifications (λ is the scaling coefficient that is discussed in detail below).

Figure 5 illustrates the dynamics of the Sos peptide bound to the surface of the wtSH3 domain. The gold spheres indicate the instantaneous position of the center of mass of the hydrophobic N-terminal segment of Sos (PPPVPP). The blue spheres correspond to the center of mass of the charged C-terminal segment (RRRR). It is apparent that the N-terminal portion of the peptide is well-localized, whereas the C-terminal portion moves a great deal. The data shown in the graph can be used to calculate the root-mean-square fluctuation for the respective mass centers, which can be conveniently expressed in a form of crystallographic *B* factors.[114] The values calculated for N- and C-terminal segments are 35 and 209 Å^2, respectively, underscoring the different dynamic status of the two regions.

The top panel in Figure 6 shows the root-mean-square deviation (rmsd) trace of the MD trajectory as calculated for the atoms of the Sos peptide using the crystal structure 1CKB as a reference. These results pertain to the ordered portion of the peptide for which the atomic coordinates are listed in the PDB file. The peptide clearly maintains the correct binding pose for the entire duration of the trajectory. For almost half of the time the rmsd remains below 2.0 Å, while the average rmsd value is just under 2.3 Å. The pattern of hydrogen bonds and salt bridges as seen in the crystal structure is also preserved in the MD simulation (Figure 6B,C). In particular, the signature hydrogen bonds P2−Y186 and P5−W169, as well as the R7−D150 salt bridge, persist throughout the simulation.

At the same time, occasional departures from the crystallographic structure are also observed. For example, at the point in time ca. 500 ns the peptide undergoes a certain amount of twist. The resulting conformation is stabilized by the P6−W169 hydrogen bond that replaces the canonical P5−W169 bond (see Figure 6B). This conformation survives for approximately 200 ns, before reverting to the original state. Another notable transition occurs at approximately 1300 ns. At this point in time, residue R8 moves away from E166 and E167 and instead

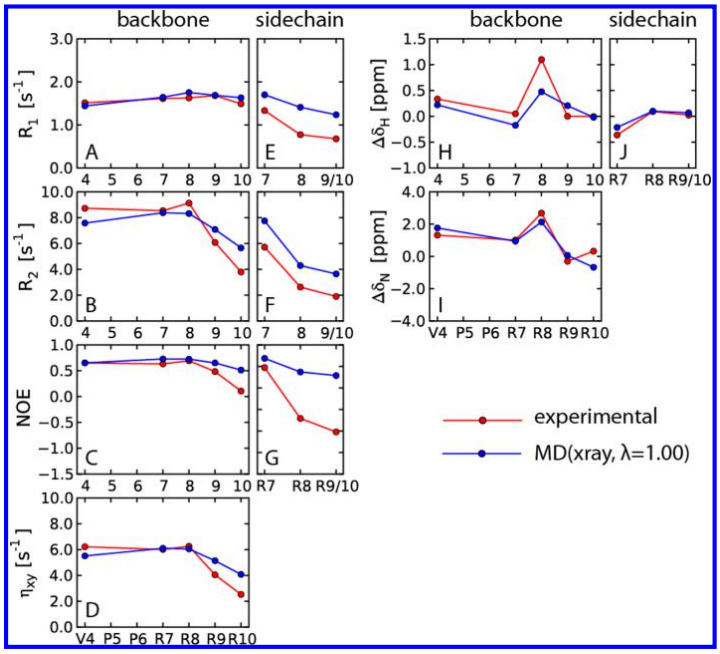


Figure 7. Comparison of the experimental (red) and simulated (blue) ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for the Sos peptide in complex with wtSH3. Simulation MD(xray,λ = 1.00) uses the original Amber ff99SB*-ILDN force field and starts from the crystallographic coordinates 1CKB. Shown are the data from the backbone amide sites in the five non-proline residues in Sos, as well as the data from four arginine side-chain ^{1}H$^{\epsilon}$,^{15}N$^{\epsilon}$ sites. The ^{1}H$^{\epsilon}$,^{15}N$^{\epsilon}$ signals from R9 and R10 are overlapped in the NMR spectra; therefore, the corresponding experimental data effectively represent an average with respect to these two residues. To allow a comparison with experiment, the simulated data from these two sites have also been averaged. SHIFTX2 software used in our work does not predict ^{15}N$^{\epsilon}$ chemical shifts; therefore, the corresponding binding shifts have not been included in the comparison. The magnitude of the experimental errors is less than or comparable to the size of the symbols. To investigate the convergence properties of MD-based calculations, we recorded an additional 1.7 μs MD trajectory of Sos:wtSH3. The results (not shown) are similar to those plotted above. However, in the case of the binding shifts, the agreement is less than perfect because the simulated values are affected by rare events, such as the rupture of the R7−D150 salt bridge (cf. Figure 6).

engages D147 and E148. At the same time, the side chain of R7 shifts in the opposite direction, i.e., away from D150 and toward E166. This rearrangement leads to an appreciable increase in the rmsd, reaching ~4 Å (see Figure 6A).

Of major interest for us is the dynamic behavior of the charged C-terminal tail of Sos, which is central to the formation of the electrostatic encounter complex. This arginine-rich tail is accommodated in the cleft between the RT loop and the n-Src loop, both of which carry stretches of negatively charged amino acids (see Figure 5B). The tail moves mostly in a sideways direction, sweeping over the surface of the protein. In doing so, it sporadically forms hydrogen bonds with wtSH3 as well as multiple salt bridges (see Figure 6B,C). Some of these interactions are stable and exist for a significant fraction of time, while others make only a brief appearance. Importantly, at each point in time, the C-terminal tail of Sos is tied to wtSH3 via a significant number of interactions (on average, 4.8 interactions, counting both salt bridges and the relevant hydrogen bonds in panels B and C of Figure 6). After a while, the tail rearranges itself, breaking some of its interactions with wtSH3 and establishing a new pattern of contacts.

As indicated above, the main focus of this study is on dynamics-modulated electrostatic interactions involving the C-terminal tail of Sos. Accordingly, the NMR data have been measured using the sample of ^{15}N-labeled Sos in the presence of unlabeled wtSH3. The sample was prepared with a 4-fold excess of wtSH3 to ensure that the spectral signals originate entirely from the bound Sos (see above). Using the standard ^{1}H,^{15}N experiments, we observed five resolved peaks from the non-proline backbone amide sites in Sos (see Figure 3A). In addition, we were also able to record the signals from four ^{1}H$^{\epsilon}$,^{15}N$^{\epsilon}$ sites in the arginine side chains, although two of the side-chain resonances are overlapped (R9 and R10). The experimental data from all these sites are plotted in Figure 7 as a function of residue number. Panels A−D and E−G of Figure 7 show ^{15}N relaxation data for the backbone and side-chain sites, respectively. Panels H−I and J show the binding shifts.

The standard set of ^{15}N relaxation data, comprised of R_1 and R_2 rates as well as ^{1}H−^{15}N saturation transfer NOEs, has been recorded as described in Materials and Methods. Initially, we reckoned with the possibility that R_2 rates may contain an exchange contribution R_{ex}, although we have not identified any exchange mechanisms that could potentially cause exchange broadening in Sos:wtSH3. From the perspective of MD modeling, the R_{ex} term represents a problem because it normally cannot be simulated (unless an ultralong trajectory is available).[115] To address this issue, we chose to measure transverse cross-correlated relaxation rates η_{xy} that carry the

same dynamic information as R_2 but are free from μs−ms exchange effects.[116,117] All relaxation data have been collected at two static magnetic field strengths, 500 and 600 MHz. For the sake of clarity, only the latter data set is illustrated in Figure 7 (the discussion of field dependence is deferred until later).

In addition, we have also quantified ^{1}H and ^{15}N peak shifts that occur in the spectrum of Sos upon binding to wtSH3 (cf. Figure 3A). A number of prominent shifts have been observed. In particular, a large downfield shift of the amide resonance from R8 (+1.1 ppm on $^1H^N$ and +2.7 ppm on ^{15}N) is due to the hydrogen bond[118] with the side-chain carboxyl group of E166. A sizable upfield shift of the signal from the R7 side chain (−0.4 ppm on $^1H^\varepsilon$ and −1.7 ppm on $^{15}N^\varepsilon$) is caused by a ring current in W169, with which R7 forms a cation−π contact. Likewise, the downfield shift of the amide signal from V4 (+0.3 ppm on $^1H^N$ and +1.3 ppm on ^{15}N) is caused by the ring current from Y186. At the same time, residues R9 and R10 show very small binding shifts; apparently, these two residues remain extended into solvent and highly disordered. Consistent with crystallographic data, two terminal arginines are only transiently involved in the interactions with wtSH3, although they contribute to the kinetics of Sos−wtSH3 association through the initial electrostatic attraction.

This array of experimental data offers a good opportunity to put our MD model to a rigorous test. There is a long history of validating MD simulations via spin relaxation rates.[106,119−121] Briefly, the MD data are used to directly compute the dipolar (as well as CSA) temporal correlation functions for the individual amide sites. These correlation functions are subsequently converted into spectral densities, which are in turn used to calculate ^{15}N relaxation parameters. Spin relaxation is directly influenced by fast internal dynamics, i.e., the motions with correlation times on the order of τ_{rot} or shorter, and indirectly by slower motional processes. For example, ^{15}N relaxation can be indirectly affected by rare conformational transitions as seen in Figure 6. Indeed, formation and dissolution of salt bridges are expected to have an effect on the fast local dynamics of Sos bound to the surface of wtSH3. From this perspective, it is important that the MD simulation is sufficiently long, 3.06 μs, to sample such relatively rare events.

To calculate binding shifts, we used the knowledge-based prediction program SHIFTX2.[90] The program was first applied to the MD trajectory representing free Sos and then to the trajectory in which Sos is bound to wtSH3, thus allowing calculation of $\Delta\delta_{Sos} = \delta_{Sos}^{bound} - \delta_{Sos}^{free}$. A number of studies have appeared recently where chemical shift predictors have been applied to the frames from the MD simulations.[110,122,123] Moreover, a new generation of predictors has been developed specifically for this purpose.[124,125] Note, however, that these previous applications have a somewhat different focus—namely, they seek to improve the prediction of chemical shifts by taking into consideration the local dynamics of proteins. In our work, the calculated shifts are used to validate the dynamic scenario seen in the MD simulation.

In particular, $\Delta\delta$ values are sensitive to the (dynamically modulated) hydrogen bonds, cation−π contacts, and salt bridges such as those formed by several residues in the C-terminal portion of Sos. Hence, the comparison of the calculated and experimental shifts should shed light on these interactions and their dynamic status (degree of motional averaging). Of note, SHIFTX2 employs well-established and reliable algorithms for calculating shifts caused by hydrogen bonds and ring currents. This makes it well-suited for the purpose of our study, i.e., to probe the interactions between Sos and wtSH3.

We have also measured binding shifts in the reverse direction, i.e., the shifts in ^{15}N-labeled wtSH3 upon binding of unlabeled Sos. The results are graphically represented in Figure S2 of the Supporting Information. As expected, the most significant chemical shift perturbations are localized in the vicinity of the binding site. In principle, it should be possible to use $\Delta\delta_{SH3}$ to further validate the MD data, along the same lines as what has been done with $\Delta\delta_{Sos}$. In practice, however, we have found that such analysis is of no value. None of the backbone amide groups in wtSH3 forms a hydrogen bond with Sos. Furthermore, Sos contains no aromatic rings. Thus, two of the largest sources of binding shifts are eliminated in this case. As a consequence, the measured $\Delta\delta_{SH3}$ values are small to moderate; none of the shifts exceeds 0.5 ppm in the $^1H^N$ dimension or 1.3 ppm in the ^{15}N dimension. Both the MD simulation and the SHIFTX2 algorithm are unable to properly capture and process the subtle structural and dynamic changes that cause shifts of this magnitude. As a result, the computed $\Delta\delta_{SH3}$ values display little or no correlation with the experimental data (r = 0.40 for $^1H^N$ shifts, and r = 0.19 for ^{15}N shifts). This cannot be construed as evidence against the MD model but rather reflects the limitations of the method.

Figure 7 presents the summary of Sos ^{15}N relaxation data and the binding shifts, as measured experimentally (red symbols) and calculated on the basis of the MD trajectory (blue symbols). Let us first focus on the backbone relaxation data from residues V4, R7, and R8. These data are reasonably well reproduced by molecular dynamics. Although certain discrepancies are seen in the R_2 plot (panel B), the agreement is nearly perfect for all other parameters, including transverse cross-correlated relaxation η_{xy} (panel D). As already mentioned, η_{xy} is one of the more accurate measures of protein dynamics that is free of the exchange effect; in terms of information content, η_{xy} is equivalent to R_2. Generally, the backbone amide sites of V4, R7, and R8 can be categorized as tightly bound. They are essentially integrated with the rigid wtSH3 scaffold; their spin relaxation is controlled by the overall tumbling motion of the complex and, to a certain degree, small local fluctuations.

On the other hand, the backbone relaxation data for residues R9 and R10 as well as side-chain data for R7, R8, R9, and R10 show clear systematic deviations between the experiment and the MD model (panels A−G). Specifically, the simulated R_2 and η_{xy} rates are higher than the corresponding experimental values. Likewise, the simulated $^1H^N$−^{15}N saturation transfer NOEs are higher than the corresponding experimental values. This clearly indicates that MD simulation underestimates the amount of local mobility for all of these sites.

According to the MD model, the C-terminal tail of Sos is attached to wtSH3 through the network of salt bridges. While this network is occasionally rearranged, it still has a strong restraining effect on the tail (see Figures 5 and 6). As a result, the relaxation properties of the tail are rather similar to those of the rest of the complex, which is well-structured and can be viewed as rigid (cf. blue profiles in panels A−G). The experimental data, however, paint a rather different picture. It appears that the pair of C-terminal residues in Sos and all the arginine side chains enjoy a degree of motional freedom much greater than that suggested by the MD simulation. This dynamic scenario is also supported by the crystallographic evidence (reviewed above). It is also worth noting that the MD simulation fails to accurately reproduce the large binding shift

of $^{1}H^{N}$ in residue R8 (panel H). This observation implicates the side chain of residue E166, which forms an on−off hydrogen bond with this amide as well as transient interactions with several other sites in Sos.

Surveying the results shown in Figure 7, we hypothesize that the root cause of the observed discrepancies is the deficiency in the MD force field, specifically, the exaggerated strength of the salt bridge interactions.[51,52] This issue is discussed at length below.

Simulations of Sos:wtSH3 Using Alternative MD Setups. Two groups of force field parameters are of principal significance for ionic (salt bridge) interactions: partial charges and Lennard-Jones parameters (see the next section). As far as partial charges are concerned, there is a considerable amount of variability between different families of force fields, as well as different generations in the same family. For instance, the charge assigned to N^{η} atoms in the arginine side chain is −0.62 in CHARMM22*, −0.80 in CHARMM27, −0.86 in Amber ff99SB, −0.69 in Amber ff03, and −0.26 in GROMOS53A6. There are also significant variations in the Lennard-Jones parameters between different force fields; these variations are largely responsible for dramatic differences in the radial distribution functions associated with salt bridge interactions.[51] A very recent study by Debiec et al. used small molecules (guanidinium and acetate) to model Arg−Asp/Glu salt bridges.[52] It has been found that six popular force fields, including CHARMM22*, CHARMM27, Amber ff99SB, Amber ff03, and the newest Amber ff13α force field, all overestimate the population of such salt bridges by a factor of roughly 2. This result essentially does not depend on the choice of water model (TIP3P, SPC/E, TIP4P-Ew, etc.).

While the results of Debiec and co-workers strongly suggest that there is a problem with parametrization of salt bridge interactions, these results are relevant for small molecules and thus may not accurately reflect the situation in peptides and proteins. Indeed, it is conceivable that in protein simulations salt bridge interactions are partially offset by the interactions between charged side chains and backbone amides. If true, this may have an effect of partial error cancellation. Therefore, it should be instructive to investigate this problem directly in the context of protein MD simulation employing different force fields.

With this goal in mind, we have tested AMBER ff03 (Amber 11 platform), CHARMM22* (Desmond platform), GROMOS53A6 (GROMACS 4.5.5 platform), and AMOEBA polarizable force field (TINKER 6.0.05 platform using the OpenMM library to support the GPU acceleration).[126−130] In each case, we have recorded MD trajectories of Sos:wtSH3 using the crystallographic structure as a starting point; the durations of the simulations were 445, 345, 108, and 19 ns, respectively. These trajectories were subsequently analyzed in the same fashion as described above. None of these trials produced the results that we had hoped for. AMBER ff03 shows less propensity to form salt bridges. However, three native hydrogen bonds connecting the backbone of Sos with side chains of wtSH3 are almost completely lost in this simulation, leading to poor predictions of chemical shifts. Similar behavior has been noted for this force field previously.[131] In the CHARMM22* trajectory, two of the native hydrogen bonds are greatly weakened, while arginine side chains remain over-restrained. Chemical shift predictions using this trajectory also prove to be poor. Finally, GROMOS 53A6 significantly underestimates the strength of electrostatic interactions: all salt bridges, including the native R7−D150 bridge, virtually disappear, whereas hydrogen bonds become transient and weak. As a consequence, the bound peptide experiences rapid dynamic fluctuations; in turn, this leads to large systematic deviations between the predicted and experimental ^{15}N relaxation rates.

Of special interest are the results from the 19 ns trajectory of Sos:wtSH3 recorded with the AMOEBA polarizable force field. The polarizability effects are deemed to be important in the context of salt bridge formation,[132−134] and polarizable force fields have demonstrated their usefulness in the context of protein−ligand binding and protein structure refinement.[135−139] Although there have been reports of the protein instability in AMOEBA simulations,[130] in our case the protein structure remains intact (within 1.5 Å of the crystal coordinates viz. the C^{α} rmsd). However, three native hydrogen bonds linking Sos and wtSH3 are absent in this simulation, and there are only two salt bridges, R7−D150 and R7−D147. Chemical shift predictions have been only partially successful, and there are indications that the C-terminal segment of Sos remains over-restrained. Given that the simulation is short on the absolute scale, these observations should not be over-interpreted; because of the high computational cost, we have not attempted to extend this AMOEBA trajectory.

In a further attempt to improve the situation, we experimented with different water models. The choice of water model is known to affect protein hydration[140−142] as well as conformational preferences and dynamics of disordered peptides.[143,144] Using the Amber ff99SB*-ILDN force field, we have tested TIP3P, SPC/E, and TIP4P-EW models.[113,145,146] In addition, we have also tested a polarizable model SWM4-NDP (supported in the GROMACS package).[147] No significant improvement has been observed in any of these trial trajectories, consistent with the recent findings.[52] We have also addressed several other details pertaining to the MD procedure. For example, phosphate ions have been included in the MD setup to reflect the presence of 20 mM phosphate in the NMR buffer. In principle, phosphate ions can form strong interactions with Arg side chains,[148] thus competing with the Sos:wtSH3 salt bridges. However, this tactic also failed to produce any improvements. Likewise, increasing the non-bonded cutoff distance from 8 to 10 Å had no effect on the simulation results. While none of these trials can be considered in any way definitive, we note that our results are consistent with the recent small molecule studies.[51,52]

Finally, we have tested the so-called MDEC (molecular dynamics in electrostatic continuum) model.[149−151] This model focuses on electrostatic interactions involving ionized side chains. In all conventional force fields, ionized side chains are assigned a net charge of +1 or −1 and their electrostatic interactions are calculated accordingly. The authors of the MDEC model argue that this approach fails to take into consideration the effect of electronic dielectric screening. To remedy this defect, the authors suggest to scale the respective point charges with the factor $\alpha = 1/(\varepsilon_{el})^{1/2}$, where the electronic screening constant for a protein in aqueous solution is estimated to be $\varepsilon_{el} = 2$. We have implemented the MDEC strategy in the Amber ff99SB force field and subsequently recorded a number of trial trajectories with different settings of α. The best results were obtained for $\alpha = 0.9$, which is significantly different from the prescribed value ($\alpha = 0.7$). The MDEC approach is admittedly empirical; a rigorous implementation of this concept would require consistent reparamet-

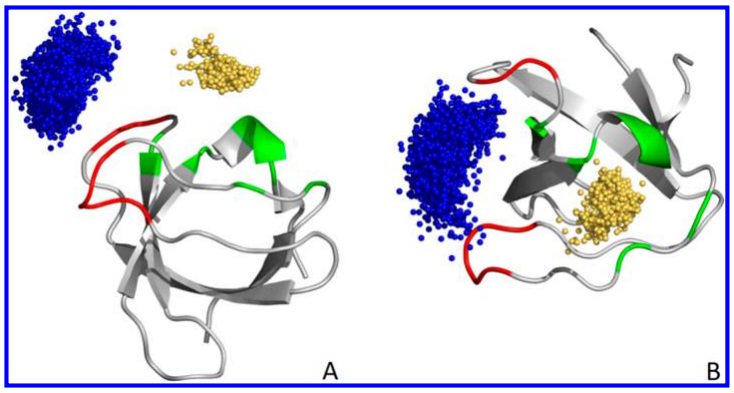


Figure 8. Position of the peptide Sos on the surface of wtSH3 according to the data from the 3.20 μs simulation using the modified force field, MD(xray,λ = 1.03). The plotting conventions are the same as in Figure 5.

rization of all force field parameters.[149] The rescaling of side-chain point charges should have an effect on many aspects of the system, including side-chain solvation, in a manner that is not easily predictable. Therefore, we abandoned this strategy in favor of a more targeted approach, as described in the next section.

Adjustment to Amber ff99SB*-ILDN. Fine-Tuning Salt Bridge Interactions. The interaction between two charged atoms is quantum mechanical in nature. When the two atoms are far apart, the interaction energy can be well approximated by the classical Coulomb's law. In contrast, when they approach each other such that the electron clouds start to overlap, the interaction becomes much more complex. In particular, there is an effect of induced polarization that gives rise to attractive dispersion forces; at short distances, there is also a strong repulsive force due to Pauli interaction. In classical molecular dynamics, all deviations from the standard Coulomb's law are empirically modeled by the Lennard-Jones (LJ) potential, eq 2. For the pair of atoms indexed i and j, the potential is defined by two parameters: the interatomic distance at which the LJ energy is at its minimum, r_0^{ij}, and the depth of the corresponding energy well, ε^{ij}. After the atomic charges are set, the LJ parameters determine how close the two atoms can approach each other and are thus strongly related to the magnitude of their optimal interaction energy.

In conventional MD force fields, the pairwise LJ parameters are normally determined by simple combination rules. In particular, each atom type is assigned a van der Waals radius, and the sum of the two radii for any two atoms is taken to be the distance r_0^{ij} for this particular pair. It has been increasingly recognized, however, that in certain cases the simple combination rules require additional fine-tuning to reproduce the desired interatomic interactions. For example, a slight increase in r_0^{ij} between Na^+ (K^+) and Cl^- (over the default value derived from the standard combination rule) made it possible to correct for the otherwise excessive ion pairing in the simulations of simple electrolytes.[152] The LJ parameters for the cation − carbonyl oxygen pairs have also been refined to better reproduce the energetics of K^+ ions in a protein channel.[153] Similarly, a small adjustment in r_0^{ij} has been used to avoid excessively strong binding of the cations to the negatively charged lipid headgroups, leading to an excellent agreement with experimental data.[154] The same approach has been employed to fine-tune the interactions between various metal ions and DNA phosphates, thus avoiding clustering artifacts in the DNA simulations.[155] Other very recent examples include interaction of Mg^{2+} with di- and triphosphate nucleotides,[156] as well as interaction between Ca^{2+} and acetate (viewed as an analogue of Asp and Glu side chains).[157] Finally, pairwise-specific LJ parameters have been implemented for aqueous solutions of alkanes, alcohols, and ethers in the CHARMM Drude polarizable force field, leading to accurate values of the corresponding hydration free energies.[158]

Here we use this philosophy to correct for the effect of overstabilized Arg−Asp/Glu salt bridges in Amber ff99SB*-ILDN simulations. Toward this end, a number of trial trajectories have been recorded where the Lennard-Jones r_0 distance for $N^{\eta}-O^{\delta/\varepsilon}$ pairs has been rescaled (see Materials and Methods). Specifically, we have tested scaling factors λ = 1.01, 1.02, 1.03, 1.04, 1.05, 1.10, 1.15, and 1.20. In each case, the MD data have been processed along the same lines as shown above (Figures 5−7). Unsurprisingly, large corrections to r_0 led to dramatic attenuation of salt bridges. A good indicator is the native salt bridge between R7 and D150, which is observed in the X-ray structure 1CKB. This salt bridge is present for 86% of the time in the original simulation, λ = 1.00 (trajectory length 3.06 μs), 70% of the time in the altered simulation with λ = 1.03 (3.2 μs), 24% of the time in the altered simulation with λ = 1.05 (0.8 μs), and is completely lost in the simulations with λ = 1.10 or higher. This result, as well as direct comparison of the simulated and experimental ^{15}N relaxation rates and binding shifts, led us to identify λ = 1.03 as the optimal scaling factor. While this is clearly an ad hoc choice, we provide an independent validation for it in paper 2, where we investigate the fuzzy complex between Sos and the double mutant of c-Crk N-SH3 (Sos·dmSH3).

It is worth noting that LJ distance parameters have been reoptimized together with partial charges in the new Amber force field ff13α. In particular, the r_0 distance for $N^{\eta}-O^{\delta/\varepsilon}$ pairs increased by 6%.[159] This new force field remains a work in progress; in its current state, Amber ff13α still shows a tendency to significantly overestimate the strength of salt bridge interactions.[52] Nevertheless, it is clear that any solution to this problem is likely to involve the Lennard-Jones parameters.

dx.doi.org/10.1021/bi500904f | *Biochemistry* 2014, 53, 6473−6495

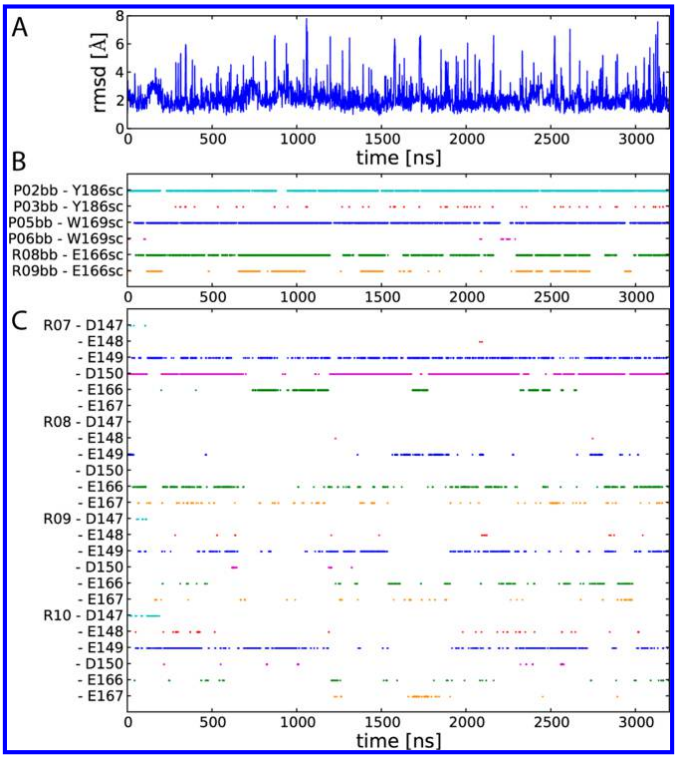


Figure 9. (A) rmsd of the peptide atomic coordinates relative to the crystallographic structure 1CKB as calculated for the 3.20 μs simulation of Sos:wtSH3 MD(xray,λ = 1.03). (B and C) Time traces representing the occurrence of peptide−protein hydrogen bonds and salt bridges in the trajectory. All conventions are the same as in Figure 6. Prominent hydrogen bonds and salt bridges are the same as those found in the crystallographic structure 1CKB (see Figure 4).

Finally, it is worth mentioning that MD simulations of proteins using implicit solvent are also faced with the problem of overstabilized salt bridges. This problem has been solved by adjusting the Born radii of the atoms that form the salt bridge.[160−162] Conceptually, this approach is similar to that described above.

Simulations of Sos:wtSH3 Using a Modified Version of Amber ff99SB*-ILDN. Comparison with Experimental Data. The calibration of Lennard-Jones parameters for N^{η}−$O^{\delta/\varepsilon}$ pairs, as discussed in the previous section, favors the distance r_0^{ij} that is 3% longer compared to the original setting. The new r_0^{ij} value has been coded into the ff99SB*-ILDN force field, thus overriding the standard combination rule for these particular off-diagonal LJ terms. Using the amended force field, we have recorded a 3.20 μs trajectory of Sos:wtSH3 beginning from the crystal coordinates. In what follows, this simulation is referred to as MD(xray,λ = 1.03) according to the choice of scaling factor λ. Aside from the change in the force field, all other details of the MD protocol were exactly the same as in the original 3.06 μs trajectory, MD(xray,λ = 1.00). The new simulation is analyzed using the same format as before; the results are shown in Figures 8−10.

Figure 8 illustrates the fluctuations of the Sos peptide when attached to the canonical binding site on the surface of wtSH3. Gold and blue spheres in the graph represent the centers of mass of the hydrophobic N-terminal segment of Sos and the arginine-rich C-terminal segment, respectively. The results are rather similar to those previously obtained from the conventional simulation (cf. Figure 5), although one can also notice certain differences. In the new trajectory, the hydrophobic portion of Sos is somewhat better localized in the binding site. The scatter of the gold spheres in Figure 8 corresponds to a temperature factor of 31 Å^2, which is slightly lower than the previously obtained value of 35 Å^2. At the same time, the arginine-rich portion of the peptide acquires a somewhat greater degree of motional freedom, 229 Å^2 vs 209 Å^2.

These observations lead us to conclude that there is a certain amount of strain in the Sos:wtSH3 complex. Specifically, formation of strong salt bridges by the C-terminal portion of Sos has a certain destabilizing effect on the binding of the N-terminal portion of Sos. Conversely, weakening of the salt bridges in the λ = 1.03 simulation creates conditions for tighter binding of the N-terminal segment. This kind of behavior is also manifested in Figures 6A and 9A. Shown in these graphs is the rmsd trace of the Sos peptide, calculated for the structured portion of Sos (those atoms for which the crystallographic coordinates are available). The graph from MD(xray,λ = 1.03) simulation clearly displays less dynamic fluctuations; in particular, there are no large-amplitude long-lived fluctuations such as those observed in the MD(xray,λ = 1.00) trajectory.

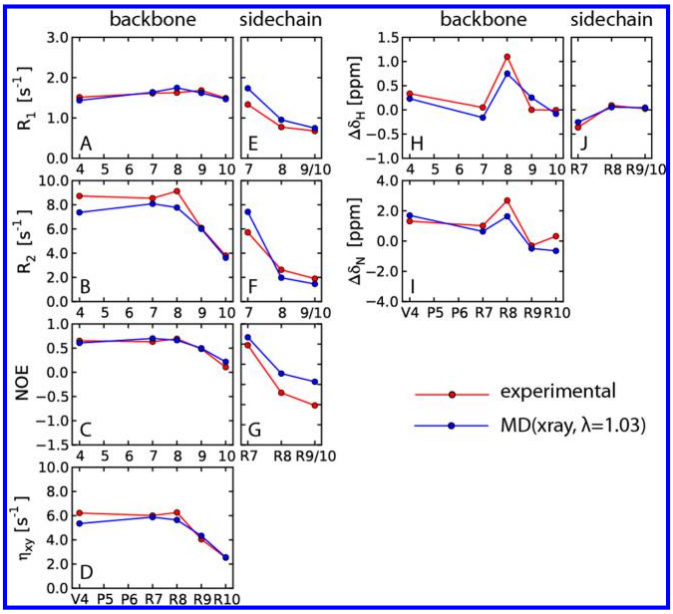


Figure 10. Comparison of the experimental (red) and simulated (blue) ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for the Sos peptide in complex with wtSH3. The MD(xray,λ = 1.03) simulation uses the modified version of the Amber ff99SB*-ILDN force field in which the strength of salt bridge interactions involving Arg and Glu/Asp side chains has been adjusted.

The average Sos rmsd in the new simulation is 2.1 Å, somewhat lower than that registered previously, 2.3 Å. Hence, reducing the strength of salt bridges appears to enhance peptide binding, which is associated primarily with the well-structured N-terminal portion of the peptide.

An interesting picture emerges from these observations. The electrostatic attraction associated with Arg−Glu/Asp pairing pulls the peptide toward its target, thus efficiently increasing the k_{on} rate. However, once the peptide is bound the same electrostatic interactions involving the C-terminal portion of Sos may no longer be conducive to binding because these interactions are not necessarily compatible with hydrophobic contacts and hydrogen bonds formed by the N-terminal portion of the peptide. In other words, the propensity to form salt bridges may cause an increase in k_u (see Figure 1), thus slightly degrading the binding affinity. This highlights the dual role of electrostatic interactions in the context of peptide binding; they are responsible for dramatic gains in the on rate, but they may also have a certain destabilizing effect on the bound state.

The traces shown in panels B and C of Figure 9 illustrate the time dependence of hydrogen bonds and salt bridges connecting Sos and wtSH3 throughout the course of the MD(xray,λ = 1.03) simulation. Obviously, the propensity of the C-terminal tail to form salt bridges is greatly diminished. On average, at any given point in time, the C-terminus forms 2.3 salt bridges. This is a 2-fold decrease compared to the original simulation, where the corresponding number was 4.3. At the same time, the hydrogen bond contacts are somewhat strengthened, 2.1 in the altered simulation versus 1.7 in the original simulation. As already discussed above, the reduced propensity to form salt bridges removes some of the steric strain, which, in turn, leads to better hydrogen bonding.

Another way of looking at it would be to determine the number of contacts formed by the structured portion of the peptide (for which the crystallographic coordinates are available) and the remaining unstructured segment. For the structured portion of Sos, the average number of contacts is 3.0 as determined from the MD(xray,λ = 1.03) trajectory. This includes the native salt bridge R7−D150, which is largely preserved in the new simulation, see Figure 9C. The result is quite similar to the one obtained from the original trajectory MD(xray,λ = 1.00), where the average number of contacts is found to be 3.1. At the same time, the data from the unstructured portion of Sos display some significant differences. The altered MD simulation predicts, on average, 1.3 contacts in this area (salt bridges as well as hydrogen bonds), whereas the original trajectory predicts 2.9 such contacts.

Similar conclusions can be drawn from the analysis of the solvent accessible surface areas. Considering the N-terminal segment of Sos, the average buried surface area is essentially the same in the two simulations, 546 and 544 Å^2. At the same time, there is a significant difference in how the C-terminal segment is packed against the protein surface—the packing is less tight in the modified simulation protocol, 439 Å^2 vs 574 Å^2.

Finally, the most important element of the validation scheme is the direct comparison between the MD-derived NMR parameters and the experimentally measured data (Figure 10). A brief survey of these results indicates a big improvement compared to the standard implementation of molecular dynamics (Figure 7). In particular, the relaxation rates of the backbone amide groups belonging to residues R9 and R10 are now in perfect agreement with the experiment. The arginine

dx.doi.org/10.1021/bi500904f | *Biochemistry* 2014, 53, 6473−6495

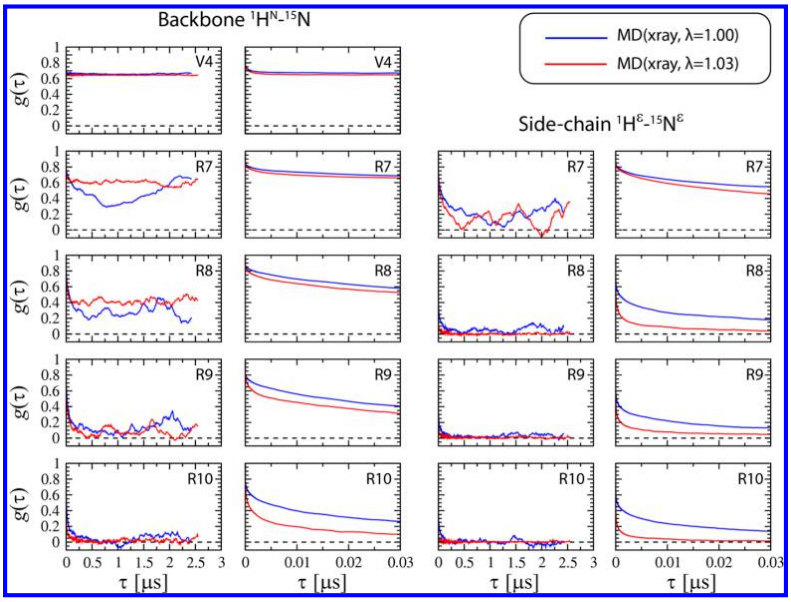


Figure 11. $^1H-^{15}N$ dipolar correlation functions $g(\tau)$ for backbone and arginine side-chain sites in Sos as extracted from MD(xray,λ = 1.00) and MD(xray,λ = 1.03) trajectories of Sos:wtSH3 (blue and red curves, respectively). The correlation functions have been calculated to 80% of the trajectory length. The expanded view of the initial portion of $g(\tau)$, which is relevant for spin relaxation, is shown in the second and fourth columns. The results reflect internal dynamics of the Sos:wtSH3 complex, with the overall reorientational motion (tumbling) factored out as described in Materials and Methods. Although it is difficult to tell with certainty, the plateau of the correlation function for backbone amides in residues R7 and R8 appears to be *lower* in the MD(xray,λ = 1.00) simulation than in the MD(xray,λ = 1.03) simulation. If this is true, this means that weakening of the salt bridges in the latter trajectory actually has a partial stabilizing effect on those two residues (similar to what has been discussed in relation to Figure 8). However, for the sake of discussion, we simply suggest here that the plateau values are similar within the uncertainty of calculations. Finally, note that the MD(xray, λ=1.03) trajectory, which features faster dynamics, shows better convergence properties than its MD(xray, λ=1.00) counterpart.

side-chain results for residues R8 and R9/R10 are also greatly improved. Furthermore, there is also a noticeable improvement in binding shifts; specifically, the large shift experienced by $^1H^N$ from residue R8 is now successfully reproduced by the MD-based calculations. At the same time, for those sites that belong to the well-structured portion of the peptide, e.g., V4, the level of agreement between the simulations and experiment remains unchanged (cf. Figures 7 and 10).

What is the structure/dynamics basis for the improvement observed in Figure 10? To answer this question, we turn to dipolar correlation functions underlying the ^{15}N relaxation rates. The full complement of these correlation functions, for both backbone and Arg side-chain sites in Sos, is shown in Figure 11. The plots in the first and third columns show the correlation functions $g(\tau)$ computed to 80% of the trajectory length. In each case, $g(\tau)$ are supposed to converge to a plateau, reflecting the long-term averaging properties of the particular $^1H-^{15}N$ vector. Although the convergence behavior is far from perfect, the data are generally sufficient to judge the amount of disorder at each site. As far as we can tell, the plateau values derived from MD(xray,λ = 1.00) and MD(xray,λ = 1.03) trajectories are similar within the uncertainty associated with incomplete convergence (cf. blue and red curves, first and third columns in Figure 11). Consider, for instance, the data for residues R9 and R10. The backbone and side-chain correlation functions in R9 and R10 essentially decay to zero. This means that in both trajectories the respective proton−nitrogen vectors experience complete orientational averaging.

The observations concerning the plateaus of the correlation functions (i.e., the order parameters) lead us to conclude that the amplitudes of internal motions in MD(xray,λ = 1.00) and MD(xray,λ = 1.03) trajectories are actually similar. This conclusion agrees with the analysis illustrated in Figures 5 and 8, where the differences between the two trajectories are relatively minor. Therefore, it is not the motional amplitudes that cause the differences in the simulated ^{15}N relaxation rates. Rather, it is the time scale of internal dynamics that plays the key role. The panels in the second and fourth columns of Figure 11 illustrate the expansion of the simulated correlation functions to 30 ns. This is the time interval that is relevant for spin relaxation in the Sos:wtSH3 complex where the tumbling time is τ_{rot} = 7 ns. From these expansion plots, it is obvious that the internal dynamics in Sos is faster in the altered MD(xray,λ = 1.03) simulation than it is in the original MD(xray,λ = 1.00) simulation. This observation is relevant for C-terminal residues, R8−R10; the effect is particularly striking for arginine side-chain sites (fourth column in Figure 11).

The analysis described above suggests the following picture of the simulated Sos dynamics. In the MD(xray,λ = 1.00) trajectory, the C-terminal segment of Sos samples different conformations, where it often becomes trapped for extended periods of time because of the overstabilized salt bridges. In

contrast, in the MD(xray,λ = 1.03) simulation, Sos tends to interconvert between different conformations more rapidly. Consequently, this model is more successful in reproducing the experimental ^{15}N relaxation rates.

Finally, valuable insight into internal dynamics can be obtained from relaxation data collected at multiple magnetic field strengths. With this goal in mind, we have measured the ^{15}N relaxation in Sos:wtSH3 also at a proton Larmor frequency of 500 MHz. The experimental results are summarized in the left column of Figure 12, where solid and dashed lines represent

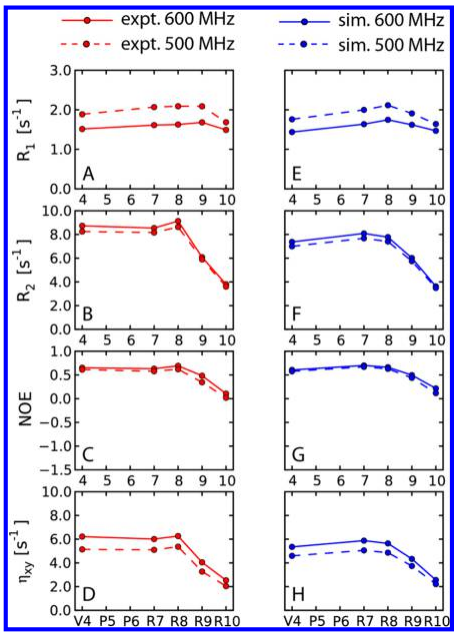


Figure 12. Experimental and simulated backbone ^{15}N relaxation rates in the peptide Sos bound to wtSH3 at 600 and 500 MHz spectrometer frequencies (solid and dashed lines, respectively). The simulated results are from the MD(xray,λ = 1.03) trajectory.

600 and 500 MHz data, respectively. The right column shows the equivalent data calculated on the basis of the MD(xray,λ = 1.03) trajectory. Insofar as the field dependence is concerned, the agreement between the experiment and the simulation is very convincing. For relatively rigid sites such as V4, the outcome is dictated mainly by the overall tumbling time, τ_{rot} = 7 ns. However, in the case of flexible residues such as R10, the results reflect a shorter correlation time, corresponding to the internal peptide dynamics (cf., for example, panels D and H in Figure 12).

Simulations of the Binding Process between Sos and wtSH3. Several years ago, Ahmad, Gu, and Helms[30] showed that it is possible to use MD simulations in explicit solvent to reproduce peptide binding in a system that is very similar to Sos:wtSH3 (the only difference is that these authors used a shorter peptide containing only two C-terminal arginines instead of four). Here we set out to reinvestigate this problem using the modified ff99SB*-ILDN force field (λ = 1.03) and significantly longer simulations. For this purpose, we have implemented a special protocol (see Materials and Methods). The initial conformation of the peptide was chosen randomly from a 2.1 μs trajectory of free Sos. The peptide was then placed at a certain distance from the surface of the protein (~15 Å considering the distance from the center of mass of Sos to the surface of wtSH3). The initial orientation of Sos was assigned randomly. The resulting system was used to construct a water box, allowing for a generous amount of water. Illustrated in Figure 13A is the initial setup that has been

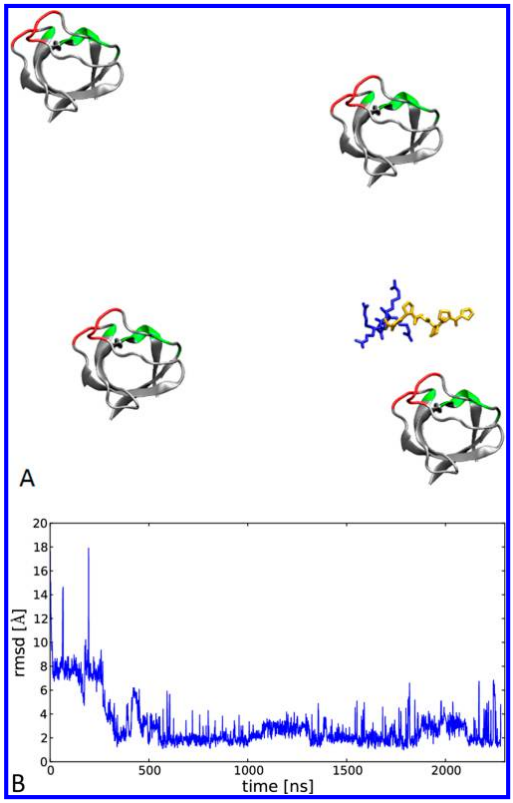


Figure 13. (A) Initial placement of the peptide and the protein in the 2.28 μs MD(rand,λ = 1.03) trajectory. Along with the actual simulated protein molecule, the plot also shows three periodic images. The space between the protein molecules is filled by TIP3P solvent. Even though the proline-rich sequence in Sos is sterically constrained, it remains sufficiently flexible in the MD simulation. (B) rmsd of the peptide atomic coordinates relative to the crystallographic structure 1CKB (same plotting conventions as in Figure 6). The lowest rmsd value is just under 1 Å. The simulation is 10-fold longer than those reported by Ahmad et al.[30]

used to record one of the MD(rand,λ = 1.03) trajectories. Along with the single simulated wtSH3 molecule, the graph also shows three periodic images of wtSH3. This representation makes it possible to appreciate the amount of water in the system (the empty space in the graph is filled with TIP3P water). In this particular trajectory, the effective concentration of the protein is 7.8 mM, which is not too far from the

dx.doi.org/10.1021/bi500904f | Biochemistry 2014, 53, 6473−6495

experimentally accessible concentrations (up to 4 mM in our experimental measurements, see the Supporting Information). Consequently, there is enough room in this simulation for the Sos peptide to fully explore the conformational space available to it, without being squeezed by the protein molecules.

Using this type of procedure, we have recorded thirty trajectories modeling the binding of Sos to wtSH3, with the trajectory length ranging from 20 to 265 ns. In nine of these trajectories the Sos peptide and wtSH3 formed a complex that was found to be similar, although not identical, to the crystal structure 1CKB. Of these nine trajectories, we have selected one that was subsequently extended to 2.28 μs. This trajectory, termed MD(rand,λ = 1.03), was used to simulate the NMR observables, which were subsequently compared with the experimental values.

In the starting frame of the MD(rand,λ = 1.03) trajectory, the Sos peptide is sufficiently far removed from the surface of wtSH3, see Figure 13. Initially, the displacement of the Sos peptide (i.e., rmsd relative to the coordinates of the bound Sos as found in the crystal structure 1CKB) amounts to 17.6 Å. As the Sos·wtSH3 encounter complex is formed, the rmsd drops rapidly to ca. 8 Å. The first salt bridge, R8−E167, is observed already after 4 ns. However, this turns out to be a transient interaction that immediately dissipates and subsequently makes only a few brief appearances throughout the trajectory. It is not before 45 ns that the two salt bridges are established in a sufficiently stable manner, R8−D150 and R8−D147. Both of these salt bridges are non-native. The resulting conformational state deserves a brief discussion. The pose of the peptide at this point generally resembles the binding pose found in 1CKB. In particular, the N-terminal portion of Sos hovers over the hydrophobic grooves where it is supposed to bind. However, the conformation of the peptide is not conducive to binding (there is a certain amount of twist); the peptide lies high above the surface of the protein and fails to form two signature hydrogen bonds, P2−Y186 and P5−W169. The hydrophobic buried surface area remains relatively small, and the peptide occasionally swings away from the surface of wtSH3 (large spikes in the initial portion of the graph in Figure 13B).

After some time, this labile state is transformed into a bona fide complex. At around 270 ns, several conformational transitions take place, proline rings P2 and P5 slide into the shallow hydrophobic grooves on the surface of wtSH3, and the native hydrogen bonds P2−Y186 and P5−W169 are established. At around 290 ns, the (weak) native salt bridge R7−E149 is formed. At 320 ns, the non-native salt bridge R8−D150 is dissolved and replaced with the native salt bridge R7−D150. As a result of these transformations, by 340 ns the coordinate rmsd of the Sos peptide drops to 1.3 Å. However, large fluctuations continue to occur until 550 ns when the complex is stabilized by the hydrogen bond between the R8 amide group and the E166 side chain, as well as certain additional salt bridges. After that, the peptide mostly remains within 2 Å of the crystallographic structure, although it still experiences substantial fluctuations involving the rearrangements of the C-terminal tail.

The MD(rand,λ = 1.03) trajectory can be used to simulate relaxation data and binding shifts that can be subsequently compared with the experimental data, in the same way as demonstrated previously. For this purpose, we have chosen the portion of the trajectory that begins at 400 ns (the point at which the Sos:wtSH3 complex is fully formed). The results are presented in Figure S3 of the Supporting Information. Briefly, they are virtually identical to those shown in Figure 10 with one exception—backbone shifts of residue R8 are significantly smaller than those observed experimentally (and those predicted on the basis of the MD(rand,λ = 1.03) trajectory). This deviation is caused by effective weakening of the hydrogen bond between R8 and the side chain of E166. In the selected time interval from 400 to 2280 ns, this hydrogen bond is identified in only 35% of all frames. By comparison, in the MD(xray,λ = 1.03) trajectory, this proportion reaches 50%. Most likely, the difference can be attributed to the less-than-perfect convergence of the MD analyses. As already indicated, 2−3 μs simulations are sufficient to reliably reproduce ^{15}N relaxation rates, but not necessarily the binding shifts. However, in principle, we cannot rule out another possibility. It is conceivable that formation of the Sos:wtSH3 complex has not been fully completed in the 2.28 μs MD(rand,λ = 1.03) simulation. Indeed, there are certain systematic differences between MD(rand,λ = 1.03) and MD(xray,λ = 1.03) trajectories—for example, the former features three intermittent salt bridges involving D147 that are virtually absent in the latter.

Finally, it is worthwhile to discuss the characteristic time scales as seen in the MD(rand,λ = 1.03) simulation. Using the values of k_{ON} and k_{OFF} experimentally determined in this study and the effective concentration of wtSH3 in the MD simulation, we can estimate that the Sos:wtSH3 complex should be formed with a time constant of ca. 100 ns and dissolve with a time constant of ca. 500 μs. These estimates are consistent with the actual simulation, in which the final complex is formed after ca. 400 ns and the peptide never becomes separated from the protein over the entire duration of the trajectory. Note that we have been successful in observing the peptide binding only due to electrostatic interactions; diffusion-limited binding is much slower and thus most likely cannot be reproduced in the relatively short MD simulations.

■ CONCLUSION

MD modeling combined with experimental NMR studies has been enormously popular in the context of (modestly sized) globular proteins. Such combined studies produced a wealth of information, greatly advancing our understanding of protein dynamics and at the same time stimulating the development of force fields. However, much of local dynamics in globular proteins is somewhat generic; e.g., small-amplitude fluctuations of peptide planes can be accurately predicted on the basis of protein structure, thus to some degree obviating the need in specialized MD studies. In contrast, intrinsically disordered proteins display a rich range of dynamic behaviors that are uniquely suited for MD and NMR studies. Yet until recently the progress in this area has been impeded by a lack of adequate computational resources. Indeed, proper modeling of an IDP requires a very large water box to accommodate the more extended protein conformations; it also requires a very long trajectory to sample the vast conformational space available to an IDP. All of this has become possible only with the advent of new computers and the suitably adapted MD algorithms.

In our study, we investigate the protein−peptide system that features a substantial element of dynamic disorder. For example, in the MD(rand,λ = 1.03) simulation, we model the flexible 10-residue peptide, which moves in a stochastic fashion toward its protein target and then forms a highly dynamic intermediate before making a transition to the bona fide complex. Furthermore, even in the bound state, the peptide

retains a substantial amount of motional freedom, especially the C-terminal residues. Although the presence of disorder in this system clearly raises the bar for MD studies (in terms of both the simulation length and the size of the water box), the simulation has been successful, reproducing the experimental NMR data with near-quantitative accuracy. We envisage that in the near future this methodology will progress from peptides to small disordered proteins and eventually to large IDPs, improving our understanding of many aspects of their function, including the disorder-to-order transition upon binding to their folded targets.

In this study, we focus on the moderately high affinity (~1 μM) complex between Sos and wild-type SH3. Comparing the experimental data (specifically, ^{15}N relaxation rates) with the results from the state-of-the-art Amber ff99SB*-ILDN simulation, we have found that molecular dynamics overemphasizes salt bridge interactions between the peptide and the protein. The tendency to misjudge the strength of salt bridges appears to be a common affliction of modern fixed-charge force fields.

To address this problem, we have devised a targeted correction involving the Lennard-Jones potential between the charged moieties of the Arg and Asp/Glu side chains. This is clearly an empirical solution; the proper way of approaching this problem would be to systematically reparameterize the entire force field. Nevertheless, there is little doubt that off-diagonal Lennard-Jones parameters associated with salt bridges need to be reassessed in Amber ff99SB as well as other widely used force fields. In this sense, we have likely identified an important target for future optimization of non-polarizable force fields.

The MD trajectory of Sos:wtSH3 recorded with the altered force field, Amber ff99SB*-ILDN (λ = 1.03), showed good agreement with the experiment, specifically, with respect to ^{15}N relaxation rates and binding shifts (in the latter case, the outcome is to some degree affected by convergence). As it turns out, weakening of the salt bridges in the C-terminal portion of Sos strengthens the binding at the N-terminal region. This implies that there is a certain amount of conflict between, on one hand, the salt bridge interactions implicated in formation of the electrostatic encounter complex and, on the other hand, hydrogen bonds and hydrophobic packing indicative of the tight complex. In other words, the interactions associated with the electrostatic encounter complex do not necessarily constitute a subset of the interactions found in the fully formed complex. As it appears, the weakening of the salt bridges in the λ = 1.03 simulation has an effect not so much on the amplitudes of motion of the C-terminal residues but on the respective correlation times. Compared to the original simulation, the C-terminal tail visits the same conformations but does not remain in these conformations for quite as long. Faster conformational exchange is reflected in the spin relaxation rates, leading to improved agreement with the experiment.

In this study, we have found the way to probe, via both MD simulation and experiment, the interactions that are key to the electrostatic encounter complex. In the second part of this work, we use the approach developed here to directly characterize the system that mimics the electrostatic encounter complex. Specifically, we have designed the double mutant of c-Crk N-SH3 (Y186L/W169F) in which the binding of the N-terminal portion of Sos is largely abrogated and the complex is held together mainly by the salt bridges, thus resembling the electrostatic encounter complex. The Sos·dmSH3 model has also been successfully characterized via NMR measurements and modeled with the help of the (modified) MD protocol. The results of this work will be described in paper 2.

■ ASSOCIATED CONTENT

Supporting Information

Table with rotational diffusion parameters for wild-type c-Crk N-SH3 in the free form and in the presence of Sos peptide, concentration dependence of chemical shifts reflecting weak self-association of wtSH3, chemical shift perturbation maps for wtSH3 self-association and for wtSH3 binding Sos, comparison of the experimental and simulated ^{15}N relaxation parameters and ^{1}H,^{15}N binding shifts for Sos:wtSH3 using the MD(rand,λ = 1.03) trajectory. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: nikolai@purdue.edu.

Present Address

[‖]Y.X.: Department of Biochemistry, Duke University, Durham, NC 27710.

Author Contributions

Y.X. and T.Y. contributed equally to this work.

Funding

This work has been supported by National Science Foundation Grant MCB 1158347.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Dr. Lake Paul for helping us set up the ITC experiment.

■ REFERENCES

(1) Williams, R. J. (1978) The conformational mobility of proteins and its functional significance. *Biochem. Soc. Trans. 6*, 1123−1126.

(2) Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Struct., Funct., Genet. 41*, 415−427.

(3) Romero, P., Obradovic, Z., Li, X. H., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins: Struct., Funct., Genet. 42*, 38−48.

(4) Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry 44*, 1989−2000.

(5) Uversky, V. N., and Dunker, A. K. (2008) Controlled Chaos. *Science 322*, 1340−1341.

(6) Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol. 337*, 635−645.

(7) Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol. 323*, 573−584.

(8) Xie, H. B., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N., and Obradovic, Z. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res. 6*, 1882−1898.

(9) Dames, S. A., Martinez-Yamout, M., De Guzman, R. N., Dyson, H. J., and Wright, P. E. (2002) Structural basis for Hif-1α/CBP recognition in the cellular hypoxic response. *Proc. Natl. Acad. Sci. U.S.A. 99*, 5271−5276.

(10) Bochkareva, E., Kaustov, L., Ayed, A., Yi, G. S., Lu, Y., Pineda-Lucena, A., Liao, J. C. C., Okorokov, A. L., Milner, J., Arrowsmith, C.

H., and Bochkarev, A. (2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc. Natl. Acad. Sci. U.S.A. 102*, 15412−15417.
(11) Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K., and Uversky, V. N. (2006) Analysis of molecular recognition features (MoRFs). *J. Mol. Biol. 362*, 1043−1059.
(12) Dancheck, B., Nairn, A. C., and Peti, W. (2008) Detailed structural characterization of unbound protein phosphatase 1 inhibitors. *Biochemistry 47*, 12346−12356.
(13) Ragusa, M. J., Dancheck, B., Critton, D. A., Nairn, A. C., Page, R., and Peti, W. (2010) Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nat. Struct. Mol. Biol. 17*, 459−464.
(14) Gunasekaran, K., Tsai, C. J., Kumar, S., Zanuy, D., and Nussinov, R. (2003) Extended disordered proteins: Targeting function with less scaffold. *Trends Biochem. Sci. 28*, 81−85.
(15) Marsh, J. A., and Forman-Kay, J. D. (2010) Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J. 98*, 2383−2390.
(16) Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., and Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res. 32*, 1037−1049.
(17) Borg, M., Mittag, T., Pawson, T., Tyers, M., Forman-Kay, J. D., and Chan, H. S. (2007) Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc. Natl. Acad. Sci. U.S.A. 104*, 9650−9655.
(18) Schreiber, G., Haran, G., and Zhou, H. X. (2009) Fundamental aspects of protein-protein association kinetics. *Chem. Rev. 109*, 839−860.
(19) Janin, J. (1997) The kinetics of protein-protein recognition. *Proteins: Struct., Funct., Genet. 28*, 153−161.
(20) Qin, S., Pang, X. D., and Zhou, H. X. (2011) Automated prediction of protein association rate constants. *Structure 19*, 1744−1751.
(21) Zhou, H. X., and Bates, P. A. (2013) Modeling protein association mechanisms and kinetics. *Curr. Opin. Struct. Biol. 23*, 887−893.
(22) Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding. *Curr. Opin. Struct. Biol. 19*, 31−38.
(23) Petros, A. M., Nettesheim, D. G., Wang, Y., Olejniczak, E. T., Meadows, R. P., Mack, J., Swift, K., Matayoshi, E. D., Zhang, H. C., Thompson, C. B., and Fesik, S. W. (2000) Rationale for Bcl-x(L)/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci. 9*, 2528−2534.
(24) Iesmantavicius, V., Dogan, J., Jemth, P., Teilum, K., and Kjaergaard, M. (2014) Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angew. Chem., Int. Ed. 53*, 1548−1551.
(25) Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L. M., Weiss, S., Hengst, L., and Kriwacki, R. W. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat. Struct. Mol. Biol. 11*, 358−364.
(26) Zhou, H. X., Pang, X. D., and Lu, C. (2012) Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Phys. Chem. Chem. Phys. 14*, 10466−10476.
(27) Wong, E. T. C., Na, D., and Gsponer, J. (2013) On the importance of polar interactions for complexes containing intrinsically disordered proteins. *PLoS Comput. Biol. 9*, e1003192.
(28) Ganguly, D., Otieno, S., Waddell, B., Iconaru, L., Kriwacki, R. W., and Chen, J. H. (2012) Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins. *J. Mol. Biol. 422*, 674−684.
(29) Ganguly, D., Zhang, W. H., and Chen, J. H. (2013) Electrostatically accelerated encounter and folding for facile recognition of intrinsically disordered proteins. *PLoS Comput. Biol. 9*, e1003363.
(30) Ahmad, M., Gu, W., and Helms, V. (2008) Mechanism of fast peptide recognition by SH3 domains. *Angew. Chem., Int. Ed. 47*, 7626−7630.
(31) Giorgino, T., Buch, I., and De Fabritiis, G. (2012) Visualizing the induced binding of SH2-phosphopeptide. *J. Chem. Theory Comput. 8*, 1171−1175.
(32) Guan, J. Y., Foerster, J. M., Drijfhout, J. W., Timmer, M., Blok, A., Ullmann, G. M., and Ubbink, M. (2014) An ensemble of rapidly interconverting orientations in electrostatic protein-peptide complexes characterized by NMR spectroscopy. *ChemBioChem 15*, 556−566.
(33) Dagliyan, O., Proctor, E. A., D'Auria, K. M., Ding, F., and Dokholyan, N. V. (2011) Structural and dynamic determinants of protein-peptide recognition. *Structure 19*, 1837−1845.
(34) Staneva, I., Huang, Y. Q., Liu, Z. R., and Wallin, S. (2012) Binding of two intrinsically disordered peptides to a multi-specific protein: A combined Monte-Carlo and molecular dynamics study. *PLoS Comput. Biol. 8*, e1002682.
(35) Matsuda, M., Tanaka, S., Nagata, S., Kojima, A., Kurata, T., and Shibuya, M. (1992) Two species of human Crk cDNA encode proteins with distinct biological activities. *Mol. Cell. Biol. 12*, 3482−3489.
(36) Reichman, C. T., Mayer, B. J., Keshav, S., and Hanafusa, H. (1992) The product of the cellular Crk gene consists primarily of SH2 and SH3 regions. *Cell Growth Differ. 3*, 451−460.
(37) Kobashigawa, Y., Sakai, M., Naito, M., Yokochi, M., Kumeta, H., Makino, Y., Ogura, K., Tanaka, S., and Inagaki, F. (2007) Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. *Nat. Struct. Mol. Biol. 14*, 503−510.
(38) Sarkar, P., Reichman, C., Saleh, T., Birge, R. B., and Kalodimos, C. G. (2007) Proline cis-trans isomerization controls autoinhibition of a signaling protein. *Mol. Cell 25*, 413−426.
(39) Camarero, J. A., Fushman, D., Sato, S., Giriat, I., Cowburn, D., Raleigh, D. P., and Muir, T. W. (2001) Rescuing a destabilized protein fold through backbone cyclization. *J. Mol. Biol. 308*, 1045−1062.
(40) Hirata, T., Nagai, H., Koizumi, K., Okino, K., Harada, A., Onda, M., Nagahata, T., Mikami, I., Hirai, K., Haraguchi, S., Jin, E. J., Kawanami, O., Shimizu, K., and Emi, M. (2004) Amplification, up-regulation and over-expression of C3G (CRK SH3 domain-binding guanine nucleotide-releasing factor) in non-small cell lung cancers. *J. Hum. Genet. 49*, 290−295.
(41) Jarzynka, N. J., Hu, B., Hui, K. M., Bar-Joseph, I., Gu, W. S., Hirose, T., Haney, L. B., Ravichandran, K. S., Nishikawa, R., and Cheng, S. Y. (2007) ELMO1 and Dock180, a bipartite Rac1 guanine nucleotide exchange factor, promote human glioma cell invasion. *Cancer Res. 67*, 7203−7211.
(42) Birge, R. B., Kalodimos, C., Inagaki, F., and Tanaka, S. (2009) Crk and CrkL adaptor proteins: Networks for physiological and pathological signaling. *Cell Commun. Signaling 7*, 7−13.
(43) Matsuda, M., Hashimoto, Y., Muroya, K., Hasegawa, H., Kurata, T., Tanaka, S., Nakamura, S., and Hattori, S. (1994) Crk protein binds to two guanine nucleotide-releasing proteins for the Ras family and modulates nerve growth factor-induced activation of Ras in PC12 cells. *Mol. Cell. Biol. 14*, 5495−5500.
(44) Feller, S. M., Knudsen, B., and Hanafusa, H. (1995) Cellular proteins binding to the first Src homology 3 (SH3) domain of the protooncogene product c-Crk indicate Crk-specific signaling pathways. *Oncogene 10*, 1465−1473.
(45) Okada, S., and Pessin, J. E. (1996) Interactions between Src homology (SH) 2/SH3 adapter proteins and the guanylnucleotide exchange factor SOS are differentially regulated by insulin and epidermal growth factor. *J. Biol. Chem. 271*, 25533−25538.
(46) Rojas, J. M., Oliva, J. L., and Santos, E. (2011) Mammalian son of sevenless guanine nucleotide exchange factors: Old concepts and new perspectives. *Genes Cancer 2*, 298−305.
(47) Wu, X. D., Knudsen, B., Feller, S. M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H., and Kuriyan, J. (1995) Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure 3*, 215−226.
(48) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and

development of improved protein backbone parameters. *Proteins 65*, 712−725.
(49) Best, R. B., and Hummer, G. (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B 113*, 9004−9015.
(50) Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins 78*, 1950−1958.
(51) Andrews, C. T., and Elcock, A. H. (2013) Molecular dynamics simulations of highly crowded amino acid solutions: Comparisons of eight different force field combinations with experiment and with each other. *J. Chem. Theory Comput. 9*, 4585−4602.
(52) Debiec, K. T., Gronenborn, A. M., and Chong, L. T. (2014) Evaluating the strength of salt bridges: A comparison of current biomolecular force fields. *J. Phys. Chem. B 118*, 6561−6569.
(53) Tolkatchev, D., Xu, P., and Ni, F. (2003) Probing the kinetic landscape of transient peptide-protein interactions by use of peptide N-15 NMR relaxation dispersion spectroscopy: Binding of an antithrombin peptide to human prothrombin. *J. Am. Chem. Soc. 125*, 12432−12442.
(54) Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature 447*, 1021−1027.
(55) Fawzi, N. L., Ying, J. F., Ghirlando, R., Torchia, D. A., and Clore, G. M. (2011) Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature 480*, 268−272.
(56) Vallurupalli, P., Bouvignies, G., and Kay, L. E. (2012) Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc. 134*, 8148−8161.
(57) Wu, K.-P., and Baum, J. (2010) Detection of transient interchain interactions in the intrinsically disordered protein α-synuclein by NMR paramagnetic relaxation enhancement. *J. Am. Chem. Soc. 132*, 5546−5547.
(58) Demers, J. P., and Mittermaier, A. (2009) Binding mechanism of an SH3 domain studied by NMR and ITC. *J. Am. Chem. Soc. 131*, 4355−4367.
(59) Terpe, K. (2003) Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol. 60*, 523−533.
(60) Huth, J. R., Bewley, C. A., Jackson, B. M., Hinnebusch, A. G., Clore, G. M., and Gronenborn, A. M. (1997) Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci. 6*, 2359−2364.
(61) Lindhout, D. A., Thiessen, A., Schieve, D., and Sykes, B. D. (2003) High-yield expression of isotopically labeled peptides for use in NMR studies. *Protein Sci. 12*, 1786−1791.
(62) Wider, G., and Dreier, L. (2006) Measuring protein concentrations by NMR spectroscopy. *J. Am. Chem. Soc. 128*, 2571−2576.
(63) Dreier, L., and Wider, G. (2006) Concentration measurements by PULCON using X-filtered or 2D NMR spectra. *Magn. Reson. Chem. 44*, S206−S212.
(64) Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry 33*, 5984−6003.
(65) Korzhnev, D. M., Skrynnikov, N. R., Millet, O., Torchia, D. A., and Kay, L. E. (2002) An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J. Am. Chem. Soc. 124*, 10743−10753.
(66) Hansen, D. F., and Kay, L. E. (2007) Improved magnetization alignment schemes for spin-lock relaxation experiments. *J. Biomol. NMR 37*, 245−255.
(67) Ferrage, F., Cowburn, D., and Ghose, R. (2009) Accurate sampling of high-frequency motions in proteins by steady-state ^{15}N-{^{1}H} nuclear Overhauser effect measurements in the presence of cross-correlated relaxation. *J. Am. Chem. Soc. 131*, 6048−6049.
(68) Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of *J* and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson. 131*, 373−378.
(69) Hall, J. B., and Fushman, D. (2003) Direct measurement of the transverse and longitudinal ^{15}N chemical shift anisotropy: Dipolar cross-correlation rate constants using ^{1}H-coupled HSQC spectra. *Magn. Reson. Chem. 41*, 837−842.
(70) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR 6*, 277−293.
(71) Goddard, T. D., and Kneller, D. G. (2002) *SPARKY 3*, University of California, San Francisco.
(72) Anafi, M., Rosen, M. K., Gish, G. D., Kay, L. E., and Pawson, T. (1996) A potential SH3 domain-binding site in the Crk SH2 domain. *J. Biol. Chem. 271*, 21365−21374.
(73) Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley Interscience, New York.
(74) Bai, Y. W., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct., Funct., Genet. 17*, 75−86.
(75) Song, J. N., Burrage, K., Yuan, Z., and Huber, T. (2006) Prediction of cis/trans isomerization in proteins using PSI-BLAST profiles and secondary structure information. *BMC Bioinf. 7*, 124.
(76) O'Neal, K. D., Chari, M. V., McDonald, C. H., Cook, R. G., YuLee, L. Y., Morrisett, J. D., and Shearer, W. T. (1996) Multiple cis-trans conformers of the prolactin receptor proline-rich motif (PRM) peptide detected by reverse-phase HPLC, CD and NMR spectroscopy. *Biochem. J. 315*, 833−844.
(77) Wu, W. J., and Raleigh, D. P. (1998) Local control of peptide conformation: Stabilization of cis proline peptide bonds by aromatic proline interactions. *Biopolymers 45*, 381−394.
(78) Kovrigin, E. L. (2012) NMR line shapes and multi-state binding equilibria. *J. Biomol. NMR 53*, 257−270.
(79) de la Torre, J. G., Huertas, M. L., and Carrasco, B. (2000) HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson., Ser. B 147*, 138−146.
(80) Lee, L. K., Rance, M., Chazin, W. J., and Palmer, A. G. (1997) Rotational diffusion anisotropy of proteins from simultaneous analysis of ^{15}N and ^{13}C$^{\alpha}$ nuclear spin relaxation. *J. Biomol. NMR 9*, 287−298.
(81) Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995) Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc. 117*, 12562−12566.
(82) Pfuhl, M., Chen, H. A., Kristensen, S. M., and Driscoll, P. C. (1999) NMR exchange broadening arising from specific low affinity protein self-association: Analysis of nitrogen-15 nuclear relaxation for rat CD2 domain 1. *J. Biomol. NMR 14*, 307−320.
(83) Bernado, P., Akerud, T., de la Torre, J. G., Akke, M., and Pons, M. (2003) Combined use of NMR relaxation measurements and hydrodynamic calculations to study protein association. Evidence for tetramers of low molecular weight protein tyrosine phosphatase in solution. *J. Am. Chem. Soc. 125*, 916−923.
(84) Noguera, V., Walker, O., Rouhier, N., Jacquot, J. P., Krimm, I., and Lancelin, J. M. (2005) NMR reveals a novel glutaredoxin-glutaredoxin interaction interface. *J. Mol. Biol. 353*, 629−641.
(85) Baryshnikova, O. K., and Sykes, B. D. (2006) Backbone dynamics of SDF-1α determined by NMR: Interpretation in the presence of monomer-dimer equilibrium. *Protein Sci. 15*, 2568−2578.
(86) Liu, Z., Zhang, W. P., Xing, Q., Ren, X. F., Liu, M. L., and Tang, C. (2012) Noncovalent dimerization of ubiquitin. *Angew. Chem., Int. Ed. 51*, 469−472.
(87) Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem. 26*, 1668−1688.
(88) Eswar, N., Eramian, D., Webb, B., Shen, M.-Y., and Sali, A. (2008) Protein structure modeling with MODELLER. In *Methods in Molecular Biology* (Kobe, B., Guss, M., and Huber, T., Eds.) pp 145−159, Humana Press, Totowa, NJ.

(89) Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very fast prediction and rationalization of pK_a values for protein-ligand complexes. *Proteins 73*, 765−783.
(90) Han, B., Liu, Y. F., Ginzinger, S. W., and Wishart, D. S. (2011) SHIFTX2: Significantly improved protein chemical shift prediction. *J. Biomol. NMR 50*, 43−57.
(91) Xue, Y., Pavlova, M. S., Ryabov, Y. E., Reif, B., and Skrynnikov, N. R. (2007) Methyl rotation barriers in proteins from ^2H relaxation data. Implications for protein structure. *J. Am. Chem. Soc. 129*, 6827−6838.
(92) Bremi, T., Brüschweiler, R., and Ernst, R. R. (1997) A protocol for the interpretation of side-chain dynamics based on NMR relaxation: Application to phenylalanines in antamanide. *J. Am. Chem. Soc. 119*, 4272−4284.
(93) Cornilescu, G., and Bax, A. (2000) Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase. *J. Am. Chem. Soc. 122*, 10143−10154.
(94) Kumar, S., and Nussinov, R. (2002) Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophys. J. 83*, 1595−1612.
(95) Jeffrey, G. A., and Saenger, W. (1991) *Hydrogen bonding in biological structures*, Springer-Verlalg, New York.
(96) Cavallo, L., Kleinjung, J., and Fraternali, F. (2003) POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res. 31*, 3364−3366.
(97) Rickles, R. J., Botfield, M. C., Weng, Z. G., Taylor, J. A., Green, O. M., Brugge, J. S., and Zoller, M. J. (1994) Identification of Src, Fyn, Lyn, Pi3k and Abl SH3 domain ligands using phage display libraries. *EMBO J. 13*, 5598−5604.
(98) Cheadle, C., Ivashchenko, Y., South, V., Searfoss, G. H., French, S., Howk, R., Ricca, G. A., and Jaye, M. (1994) Identification of a Src SH3 domain binding motif by screening a random phage display library. *J. Biol. Chem. 269*, 24034−24039.
(99) Knudsen, B. S., Zheng, J., Feller, S. M., Mayer, J. P., Burrell, S. K., Cowburn, D., and Hanafusa, H. (1995) Affinity and specificity requirements for the first Src homology 3 domain of the Crk proteins. *EMBO J. 14*, 2191−2198.
(100) Schreiber, G., and Fersht, A. R. (1996) Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol. 3*, 427−431.
(101) Wafer, L. N., Streicher, W. W., McCallum, S. A., and Makhatadze, G. I. (2012) Thermodynamic and kinetic analysis of peptides derived from CapZ, NDR, p53, HDM2, and HDM4 binding to human S100B. *Biochemistry 51*, 7189−7201.
(102) Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., and Hallewell, R. A. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature 358*, 347−351.
(103) Northrup, S. H., and Erickson, H. P. (1992) Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. U.S.A. 89*, 3338−3342.
(104) Frisch, C., Fersht, A. R., and Schreiber, G. (2001) Experimental assignment of the structure of the transition state for the association of barnase and barstar. *J. Mol. Biol. 308*, 69−77.
(105) Arai, M., Ferreon, J. C., and Wright, P. E. (2012) Quantitative analysis of multisite protein-ligand interactions by NMR: Binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J. Am. Chem. Soc. 134*, 3792−3803.
(106) Showalter, S. A., and Bruschweiler, R. (2007) Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field. *J. Chem. Theory Comput. 3*, 961−975.
(107) Penev, E., Ireta, J., and Shea, J. E. (2008) Energetics of infinite homopolypeptide chains: A new look at commonly used force fields. *J. Phys. Chem. B 112*, 6872−6877.
(108) Aliev, A. E., and Courtier-Murias, D. (2010) Experimental verification of force fields for molecular dynamics simulations using Gly-Pro-Gly-Gly. *J. Phys. Chem. B 114*, 12358−12375.
(109) Cerutti, D. S., Freddolino, P. L., Duke, R. E., and Case, D. A. (2010) Simulations of a protein crystal with a high resolution X-ray structure: Evaluation of force fields and water models. *J. Phys. Chem. B 114*, 12811−12824.
(110) Li, D. W., and Bruschweiler, R. (2010) Certification of molecular dynamics trajectories with NMR chemical shifts. *J. Phys. Chem. Lett. 1*, 246−248.
(111) Beauchamp, K. A., Lin, Y. S., Das, R., and Pande, V. S. (2012) Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput. 8*, 1409−1414.
(112) Piana, S., Lindorff-Larsen, K., and Shaw, D. E. (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophys. J. 100*, L47−L49.
(113) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926−935.
(114) Chandrasekhar, I., Clore, G. M., Szabo, A., Gronenborn, A. M., and Brooks, B. R. (1992) A 500-ps molecular dynamics simulation study of interleukin-1β in water: Correlation with nuclear magnetic resonance spectroscopy and crystallography. *J. Mol. Biol. 226*, 239−250.
(115) Xue, Y., Ward, J. M., Yuwen, T. R., Podkorytov, I. S., and Skrynnikov, N. R. (2012) Microsecond time-scale conformational exchange in proteins: Using long molecular dynamics trajectory to simulate NMR relaxation dispersion data. *J. Am. Chem. Soc. 134*, 2555−2562.
(116) Kroenke, C. D., Loria, J. P., Lee, L. K., Rance, M., and Palmer, A. G. (1998) Longitudinal and transverse ^1H-^{15}N dipolar ^{15}N chemical shift anisotropy relaxation interference: Unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *J. Am. Chem. Soc. 120*, 7905−7915.
(117) Hall, J. B., and Fushman, D. (2003) Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J. Biomol. NMR 27*, 261−275.
(118) Shen, Y., and Bax, A. (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR 38*, 289−302.
(119) Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M., and Ernst, R. R. (1992) Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution. *J. Am. Chem. Soc. 114*, 2289−2302.
(120) Pfeiffer, S., Fushman, D., and Cowburn, D. (2001) Simulated and NMR-derived backbone dynamics of a protein with significant flexibility: A comparison of spectral densities for the β ARK PH domain. *J. Am. Chem. Soc. 123*, 3021−3036.
(121) Nederveen, A. J., and Bonvin, A. M. J. J. (2005) NMR relaxation and internal dynamics of ubiquitin from a 0.2 μs MD simulation. *J. Chem. Theory Comput. 1*, 363−374.
(122) Markwick, P. R. L., Cervantes, C. F., Abel, B. L., Komives, E. A., Blackledge, M., and McCammon, J. A. (2010) Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. *J. Am. Chem. Soc. 132*, 1220−1221.
(123) Robustelli, P., Stafford, K. A., and Palmer, A. G. (2012) Interpreting protein structural dynamics from NMR chemical shifts. *J. Am. Chem. Soc. 134*, 6365−6374.
(124) Lehtivarjo, J., Tuppurainen, K., Hassinen, T., Laatikainen, R., and Perakyla, M. (2012) Combining NMR ensembles and molecular dynamics simulations provides more realistic models of protein structures in solution and leads to better chemical shift prediction. *J. Biomol. NMR 52*, 257−267.
(125) Li, D. W., and Bruschweiler, R. (2012) PPM: A side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. *J. Biomol. NMR 54*, 257−265.
(126) Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G. M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. M., and Kollman, P. (2003) A point-charge force field for molecular

mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem. 24*, 1999−2012.
(127) MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B 102*, 3586−3616.
(128) Mackerell, A. D., Feig, M., and Brooks, C. L. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem. 25*, 1400−1415.
(129) Oostenbrink, C., Villa, A., Mark, A. E., and Van Gunsteren, W. F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem. 25*, 1656−1676.
(130) Ponder, J. W., Wu, C. J., Ren, P. Y., Pande, V. S., Chodera, J. D., Schnieders, M. J., Haque, I., Mobley, D. L., Lambrecht, D. S., DiStasio, R. A., Head-Gordon, M., Clark, G. N. I., Johnson, M. E., and Head-Gordon, T. (2010) Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B 114*, 2549−2564.
(131) Trbovic, N., Kim, B., Friesner, R. A., and Palmer, A. G. (2008) Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. *Proteins 71*, 684−694.
(132) Frigyes, D., Alber, F., Pongor, S., and Carloni, P. (2001) Arginine-phosphate salt bridges in protein-DNA complexes: A Car-Parrinello study. *J. Mol. Struct.: THEOCHEM 574*, 39−45.
(133) Jiao, D., Golubkov, P. A., Darden, T. A., and Ren, P. (2008) Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U.S.A. 105*, 6290−6295.
(134) Ng, A. H., and Snow, C. D. (2011) Polarizable protein packing. *J. Comput. Chem. 32*, 1334−1344.
(135) Liang, T., and Walsh, T. R. (2006) Molecular dynamics simulations of peptide carboxylate hydration. *Phys. Chem. Chem. Phys. 8*, 4410−4419.
(136) Kaminsky, J., and Jensen, F. (2007) Force field modeling of amino acid conformational energies. *J. Chem. Theory Comput. 3*, 1774−1788.
(137) Jiao, D., Zhang, J. J., Duke, R. E., Li, G. H., Schnieders, M. J., and Ren, P. Y. (2009) Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J. Comput. Chem. 30*, 1701−1711.
(138) Fenn, T. D., Schnieders, M. J., Brunger, A. T., and Pande, V. S. (2010) Polarizable atomic multipole X-ray refinement: Hydration geometry and application to macromolecules. *Biophys. J. 98*, 2984−2992.
(139) Jiang, J. L., Wu, Y. B., Wang, Z. X., and Wu, C. (2010) Assessing the performance of popular quantum mechanics and molecular mechanics methods and revealing the sequence-dependent energetic features using 100 tetrapeptide models. *J. Chem. Theory Comput. 6*, 1199−1209.
(140) Wong, S. E., Bernacki, K., and Jacobson, M. (2005) Competition between intramolecular hydrogen bonds and solvation in phosphorylated peptides: Simulations with explicit and implicit solvent. *J. Phys. Chem. B 109*, 5249−5258.
(141) Hess, B., and van der Vegt, N. F. A. (2006) Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models. *J. Phys. Chem. B 110*, 17616−17626.
(142) Paschek, D., Day, R., and Garcia, A. E. (2011) Influence of water-protein hydrogen bonding on the stability of Trp-cage miniprotein. A comparison between the TIP3P and TIP4P-Ew water models. *Phys. Chem. Chem. Phys. 13*, 19840−19847.
(143) Florova, P., Sklenovsky, P., Banas, P., and Otyepka, M. (2010) Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact. *J. Chem. Theory Comput. 6*, 3569−3579.
(144) Nerenberg, P. S., and Head-Gordon, T. (2011) Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides. *J. Chem. Theory Comput. 7*, 1220−1230.
(145) Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987) The missing term in effective pair potentials. *J. Phys. Chem. 91*, 6269−6271.
(146) Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L., and Head-Gordon, T. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys. 120*, 9665−9678.
(147) Lamoureux, G., Harder, E., Vorobyov, I. V., Roux, B., and MacKerell, A. D. (2006) A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett. 418*, 245−249.
(148) Woods, A. S., and Ferre, S. (2005) Amazing stability of the arginine-phosphate electrostatic interaction. *J. Proteome Res. 4*, 1397−1402.
(149) Leontyev, I., and Stuchebrukhov, A. (2011) Accounting for electronic polarization in non-polarizable force fields. *Phys. Chem. Chem. Phys. 13*, 2613−2626.
(150) Leontyev, I. V., and Stuchebrukhov, A. A. (2010) Electronic continuum model for molecular dynamics simulations of biological molecules. *J. Chem. Theory Comput. 6*, 1498−1508.
(151) Zhu, K., Shirts, M. R., and Friesner, R. A. (2007) Improved methods for side chain and loop predictions via the protein local optimization program: Variable dielectric model for implicitly improving the treatment of polarization effects. *J. Chem. Theory Comput. 3*, 2108−2119.
(152) Luo, Y., and Roux, B. (2010) Simulation of osmotic pressure in concentrated aqueous salt solutions. *J. Phys. Chem. Lett. 1*, 183−189.
(153) Berneche, S., and Roux, B. (2001) Energetics of ion conduction through the K^+ channel. *Nature 414*, 73−77.
(154) Venable, R. M., Luo, Y., Gawrisch, K., Roux, B., and Pastor, R. W. (2013) Simulations of anionic lipid membranes: Development of interaction-specific ion parameters and validation using NMR data. *J. Phys. Chem. B 117*, 10183−10192.
(155) Yoo, J. J., and Aksimentiev, A. (2012) Improved parametrization of Li^+, Na^+, K^+, and Mg^{2+} ions for all-atom molecular dynamics simulations of nucleic acid systems. *J. Phys. Chem. Lett. 3*, 45−50.
(156) Simonson, T., and Satpati, P. (2013) Simulating GTP:Mg and GDP:Mg with a simple force field: A structural and thermodynamic analysis. *J. Comput. Chem. 34*, 836−846.
(157) Kahlen, J., Salimi, L., Sulpizi, M., Peter, C., and Donadio, D. (2014) Interaction of charged amino-acid side chains with ions: An optimization strategy for classical force fields. *J. Phys. Chem. B 118*, 3960−3972.
(158) Baker, C. M., Lopes, P. E. M., Zhu, X., Roux, B., and MacKerell, A. D. (2010) Accurate calculation of hydration free energies using pair-specific Lennard-Jones parameters in the CHARMM Drude polarizable force field. *J. Chem. Theory Comput. 6*, 1181−1198.
(159) Cerutti, D. S., Rice, J. E., Swope, W. C., and Case, D. A. (2013) Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. *J. Phys. Chem. B 117*, 2328−2338.
(160) Geney, R., Layten, M., Gomperts, R., Hornak, V., and Simmerling, C. (2006) Investigation of salt bridge stability in a generalized born solvent model. *J. Chem. Theory Comput. 2*, 115−127.
(161) Okur, A., Wickstrom, L., and Simmerling, C. (2008) Evaluation of salt bridge structure and energetics in peptides using explicit, implicit, and hybrid solvation models. *J. Chem. Theory Comput. 4*, 488−498.
(162) Shang, Y., Nguyen, H., Wickstrom, L., Okur, A., and Simmerling, C. (2011) Improving the description of salt bridge strength and geometry in a generalized Born model. *J. Mol. Graphics Modell. 29*, 676−684.